

ATGGGTGCGAGAGCGTCAGTATTAAGCBBBBBAGAATTAGATCGATGGGAAAAAAT  
TCGGTTAAGGCCAGGGGGAAAGAAGAAGTACAAGCTAACGCACATCGTATGGCAA  
GCAGGGAGCTAGAACGATTGCAGTTAACCTGGCTGTTAGAACATCAGAAGGC  
TGTAGACAAATACTGGGACAGCTACAACCATCCCTCAGACAGGATCAGAGGAGCT  
TCGATCACTATAACACAGTAGCAACCCCTATTGTGTGCACCAGCGGATCGAGA  
TCAAGGACACCAAGGAAGCTTAGACAAGATAGAGGAAGAGCAAAACAAGTCCAAG  
AAGAAGGCCAGCAGCAGCTGACACAGGACACAGCAATCAGGTAGCCAAA  
TTACCCCTATAGTGCAGAACATCCAGGGCAAATGGTACATCAGGCCATATCACCTA  
GAACCTTAAATGCATGGTAAAAGTAGTAGAAGAGAAGGCTTCAGCCCAGAAGTG  
ATACCCATGTTTCAGCATTATCAGAAGGAGCCACCCACAGGACCTGAACACGAT  
GTTGAACACCGTGGGGGACATCAAGCAGCCATGCAAATGTTAAAGAGACCATCA  
ATGAGGAAGCTGCAGAATGGATAGAGTCATCCAGTGCATGCAGGGCTATTGCA  
CCAGGCCAGATGAGAGAACCAAGGGAAAGTGACATAGCAGGAACACTAGTACCT  
TCAGGAACAAATAGGATGGATGACAAATAATCCACCTATCCCAGTAGGAGAGATCT  
ACAAGAGGTGGATAATCCTGGATTGAAACAAGATCGTGAGGATGTATAGCCCTACC  
AGCATTCTGGACATAAGACAAGGACCAAGGAACCCTTAGAGACTATGTAGACCG  
GTTCTATAAAACTCTAAGAGCTGAGCAAGCTTCACAGGAGGTAAAAATTGGATGA  
CAGAAACCTTGGTCCAAATGCGAACCCAGATTGTAAGACCATCCTGAAGGCT  
CTCGGCCAGCGGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTAGGAGG  
ACCCGGCCATAAGGCAAGAGTTTGGCGAGGCGATGAGCCAGGTGACGAACCTGG  
CGACCATAATGATGCAGAGAGGCAACTCCGGAACCAGCGGAAGATCGTCAAGTGC  
TTCAATTGTGGCAAAGAAGGGCACACCGCCAGGAACGTGCCGGCCCCCGGAAGAA  
GGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGAC

FIG. I

AGGCTAATTTTAAAGGAAGATCTGGCCTTCCCTACAAGGGAAGGCCAGGGAAATTT  
CTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAAGAGAGCTCAGGTCTGGGGT  
AGAGACAACAACCTCCCCCTCAGAAGCAGGAGCCGATAGACAAGGAACGTATCCTT  
TAACCTCCCTCAGATCACTCTTGGCAACGACCCCTCGTCACAGTAAGGATCGGGG  
GGCAACTCAAGGAAGCGCTGCTCGATACAGGAGCAGATGATACTAGTATTAGAAGAA  
ATGAGTTGCCAGGAAGATGGAAACCAAAATGATAGGGGGATCGGGGGCTTCAT  
CAAGGTGAGGCAGTACGACCAGATACTCATAGAAATCTGGACATAAAGCTATAG  
GTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACC  
CAGATCGGCTGCACCTTGAACCTCCCCATCAGCCCTATTGAGACGGTGCCCGTGAA  
GTTGAAGCCGGGATGGACGGCCCAAGGTCAAGCAATGGCATTGACGAAAGAGA  
AGATCAAGGCCTAGTCGAAATCTGTACAGAGATGGAGAAGGAAGGGAAAGATCAGC  
AAGATCGGGCCTGAGAACCCCTACAACACTCCAGTCTCGCAATCAAGAAGAAGGA  
CAGTACCAAGTGGAGAAAGCTGGTGGACTTCAGAGAGCTGAACAAAGAGAACTCAGG  
ACTTCTGGGAAGTCAGCTGGCATCCCACATCCCGCTGGTTGAAGAAGAAGAAG  
TCAGTGACAGTGCTGGATGGGTGATGCCACTTCTCCGTTCCCTGGACGAGGA  
CTTCAGGAAGTACACTGCCTTCACGATACTAGCATCAACAAACGAGACACCAGGCA  
TCCGCTACCAAGTACAACGTGCTGCCACAGGGATGGAAGGGATCACCAGCCATCTT  
CAAAGCAGCATGACCAAGATCCTGGAGCCCTCCGCAAGCAAAACCCAGACATCGT  
GATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAGATCGGGCAGC  
ACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGGACTGACCACA  
CCAGACAAGAACGACCAAGAAGGAACCTCCCTCGATGGGCTACGAACGTGCA  
TCCTGACAAGTGGACAGTGCAGCCATCGTGCTGCCCTGAGAAGGACAGCTGGACTG  
TGAACGACATACAGAAGCTCGTGGCAAGTTGAACCTGGCAAGCCAGATCTACCCA  
GGCATCAAAGTTAGGCAGCTGTGCAAGCTGCTCGAGGAACCAAGGCAGTACAGA

AGTGATCCCACGTACAGAGGAAGCAGAGCTAGAACTGGCAGAGAACCGAGAGATCC  
TGAAGGAGCCAGTACATGGAGTGTACTACGACCCAAGCAAGGACCTGATCGCAGAG  
ATCCAGAAGCAGGGCAAGGCCAATGGACCTACCAAATCTACCAGGAGCCCTCAA  
GAACCTGAAGACAGGCAAGTACGCAAGGATGAGGGTGCCACACCAACGATGTGA  
AGCAGCTGACAGAGGCAGTGCAGAAGATCACCACAGAGAGCATCGTATCTGGGC  
AAGACTCCAAGTTCAAGCTGCCATACAGAAGGAGACATGGGAGACATGGTGGAC  
CGAGTACTGGCAAGCCACCTGGATCCCTGAGTGGAGTTCGTGAACACCCCTCCCT  
TGGTGAAACTGTGGTATCAGCTGGAGAAGGAACCCATCGTGGGAGCAGAGACCTTC  
TACGTGGATGGGCAGCCAACAGGGAGACCAAGCTGGCAAGGCAGGCTACGTGAC  
CAACCGAGGACGACAGAAAGTGGTACCCCTGACTGACACCACCAACCAGAAAGACTG  
AGCTGCAAGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGAACATCGT  
ACAGACTCACAGTACGCACTGGCATCATCCAAGCACAACCAGACCAATCCGAGTC  
AGAGCTGGTGAACCAGATCATCGAGCAGCTGATCAAGAAGGAGAAAGTGTACCTGG  
CATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAATTAGTC  
AGTGCTGGATCCGGAAAGGTGCTGTTCTGGACGGATCGATAAGGCCAAGATGA  
ACATGAGAAGTACCACTCCAACCTGGCGCGCTATGCCAGCGACTTCAACCTGCCAC  
CTGTAGTAGCAAAGAAATAGTAGCCAGCTGTGATAAATGTCAGCTAAAAGGAGAA  
GCCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAGCTGGACTGCACGCA  
CCTGGAGGGAAAGGTGATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATAGAAG  
CAGAAGTTATCCCTGCTGAAACTGGCAGGAAACAGCATATTTCTTTAAAATTA  
GCAGGAAGATGCCAGTAAAAACAATACACACGGACAACGGAAGCAACTCACTGG  
TGCTACGGTTAAGGCCGCTGTTGGTGGCGGAAATCAAGCAGGAATTGGAATT  
CCTACAATCCCCATCGCAAGGAGTCGTGGAGAGCATGAACAAGGAGCTGAAGAAG  
ATCATCGGACAAGTGAGGGATCAGGCTGAGCACCTGAAGACAGCAGTGCAGATGGC

AGTGTTCATCCACAACTTCAAAAGAAAAGGGGGGATTGGGGGTACAGTGCAGGGG  
AAAGGATCGTGGACATCATGCCACCGACATCCAACCAAGGAGCTGCAGAAGCAG  
ATCACCAAGATCCAGAACCTCCGGGTGTACTACCGCGACAGCCGCAACCCACTGTG  
GAAGGGACCAGCAAAGCTCCTCTGGAAGGGAGAGGGGGCAGTGGTGTCCAGGACA  
ACAGTGACATCAAAGTGGTGCCAAGGCGCAAGGCCAAGATCATCCGCGACTATGGA  
AAACAGATGGCAGGTGATGATTGTGGCAAGTAGACAGGATGAGGATTAGAACCT  
GGAAGAGCCTGGTGAAGCACCATATG (SEQUENCE ID NO:1)

>wildtype	TGTACAGAGA TGGAAAAGGA AGGGAAAATT TCAAAAATTG
>mutated	TGTACAGAGA TGGAGAAGGA AGGGAAAGATC AGCAAGATCG
#1	.....*
>wildtype	GGCCTGAAAA TCCATACAAT ACTCCAGTAT TTGCCATAAA
>mutated	GGCCTGAGAA CCCCTACAAC ACTCCAGTCT TCGCAATCAA
#41	.....*
>wildtype	GAAAAAAAGAC AGTACTAAAT GGAGAAAATT AGTAGATTTC
>mutated	GAAGAAGGAC AGTACCAAGT GGAGAAAAGCT GGTGGACTTC
#81	.....*
>wildtype	AGAGAACTTA ATAAGAGAAC TCAGAGACTTC TGGGAAGTTTC
>mutated	AGAGAGCTGA ACAAGAGAAC TCAGGAGACTTC TGGGAAGTTTC
#121	.....*
>wildtype	AATTAGGAAT ACCACATCCC GCAGGGTTAA AAAAGAAAAAA
>mutated	AGCTGGGCAT CCCACATCCC GCTGGGTGA AGAAGAAGAA
#161	.....*
>wildtype	ATCAGTAACA GTACTGGATG TGGGTGATGC ATATTTTCA
>mutated	GTCAGTGACA GTGCTGGATG TGGGTGATGC CTACTTCTCC
#201	.....*
>wildtype	GTTCCTCTAG ATGAAGACTT CAGGAATAT ACTGCATTAA
>mutated	GTTCCTCTGG ACGAGGACTT CAGGAAGTAC ACTGCCTTCA
#241	.....*
>wildtype	CCATACCTAG TATAAACAAAT GAGACACCAG GGATTAGATA
>mutated	CGATACCTAG CATCAACAAAC GAGACACCAG GCATCCGCTA
#281	.....*
>wildtype	TCAGTACAAT GTGCTTCCAC AGGGATGGAA AGGATCACCA
>mutated	CCAGTACAAC GTGCTGCCAC AGGGATGGAA GGGATCACCA
#321	.....*
>wildtype	GCAATATTCC AAAGTAGCAT GACAAAATC TTAGAGCCTT
>mutated	GCCATCTTTC AAAGCAGCAT GACCAAGATC CTGGAGCCCT
#361	.....*
>wildtype	TTAGAAAACA AAATCCAGAC ATAGTTATCT ATCAATACAT
>mutated	TCCGCAAGCA AAACCCAGAC ATCGTGATCT ATCAGTACAT
#401	.....*

FIG. 2

>wildtype	GGATGATTG TATGTAGGAT CTGACTTAGA AATAGGGCAG
>mutated	GGACGACCTC TACGTAGGAA GTGACCTGGA GATCGGGCAG
#441	.....*
>wildtype	CATAGAACAA AAATAGAGGA GCTGAGACAA CATCTGTTGA
>mutated	CACAGGACCA AGATCGAGGA GCTGAGACAG CATCTGTTGA
#481	.....*
>wildtype	GGTGGGGACT TACCACACCA GACAAAAAAC ATCAGAAAGA
>mutated	GGTGGGGACT GACCACACCA GACAAGAAC ACCAGAAGGA
#521	.....*
>wildtype	ACCTCCATTG CTTTGGATGG GTTATGAAC CCATCCTGAT
>mutated	ACCTCCCTTC CTGTGGATGG GCTACGAAC GCATCCTGAC
#561	.....*
>wildtype	AAATGGACAG TACAGCCTAT AGTGCTGCCA GAAAAGACA
>mutated	AAGTGGACAG TGCAGCCCCAT CGTGCTGCCT GAGAAGGACA
#601	.....*
>wildtype	GCTGGACTGT CAATGACATA CAGAAGTTAG TGGGGAAATT
>mutated	GCTGGACTGT GAACGACATA CAGAAGCTCG TGGGCAAGTT
#641	.....*
>wildtype	GAATTGGGCA AGTCAGATTT ACCCAGGGAT TAAAGTAAGG
>mutated	GAACTGGGCA AGCCAGATCT ACCCAGGCAT CAAAGTTAGG
#681	.....*
>wildtype	CAATTATGTA AACTCCTTAG AGGAACCAAA GCACTAACAG
>mutated	CAGCTGTGCA AGCTGCTTCG AGGAACCAAG GCACTGACAG
#721	.....*
>wildtype	AAGTAATACC ACTAACAGAA GAAGCAGAGC TAGAACTGGC
>mutated	AAGTGATCCC ACTGACAGAG GAAGCAGAGC TAGAACTGGC
#761	.....*
>wildtype	AGAAAACAGA GAGATTCTAA AAAAACCAAGT ACATGGAGTG
>mutated	AGAGAACCGA GAGATCCTGA AGGAGCCAGT ACATGGAGTG
#801	.....*
>wildtype	TATTATGACC CATCAAAGA CTTAATAGCA GAAATACAGA
>mutated	TACTACGACC CAAGCAAGGA CCTGATCGCA GAGATCCAGA
#841	.....*

>wildtype	AGCAGGGGCA AGGCCAATGG ACATATCAA TTTATCAAGA
>mutated	AGCAGGGGCA AGGCCAATGG ACCTACCAAA TCTACCAGGA
#881	.....
>wildtype	GCCATTAAA AATCTGAAAA CAGGAAAATA TGCAAGAATG
>mutated	GCCCTTCAAG AACCTGAAGA CAGGCAAGTA CGCAAGGATG
#921	.....
>wildtype	AGGGGTGCC C ACACTAATGA TGAAAACAA TTAACAGAGG
>mutated	AGGGGTGCC C ACACCAACGA TGTGAAGCAG CTGACAGAGG
#961	.....
>wildtype	CAGTGCAAAA AATAACCACA GAAAGCATAG TAATATGGGG
>mutated	CAGTGCAAGAA GATCACCACA GAGAGCATCG TGATCTGGGG
#1001	.....
>wildtype	AAAGACTCCT AAATTTAAC TGCCCATA CA AAAGGAAACA
>mutated	CAAGACTCCC AAGTTCAAGC TGCCCATA CA GAAGGAGACA
#1041	.....
>wildtype	TGGGAAACAT GGTGGACAGA GTATTGGCAA GCCACCTGG
>mutated	TGGGAGACAT GGTGGACCGA GTACTGGCAA GCCACCTGG
#1081	.....
>wildtype	TTCCTGAGTG GGAGTTGTT AATACCCCTC CTTTAGTGAA
>mutated	TCCCTGAGTG GGAGTTCGTG AACACCCCTC CCTTGGTGAA
#1121	.....
>wildtype	ATTATGGTAC CAGTTAGAGA AAGAACCCAT AGTAGGAGCA
>mutated	ACTGTGGTAT CAGCTGGAGA AGGAACCCAT CGTGGGAGCA
#1161	.....
>wildtype	GAAACCTTCT ATGTAGATGG GGCAGCTAAC AGGGAGACTA
>mutated	GAGACCTTCT ACGTGGATGG GGCAGCCAAC AGGGAGACCA
#1201	.....
>wildtype	AATTAGGAAA AGCAGGATAT GTTACTAATA GAGGAAGACA
>mutated	AGCTGGCAA GGCAGGCTAC GTGACCAACC GAGGACGACA
#1241	.....
>wildtype	AAAAGTTGTC ACCCTAACTG ACACAACAAA TCAGAAGACT
>mutated	GAAAGTGGTG ACCCTGACTG ACACCCACCA CCAGAAGACT
#1281	.....

>wildtype >mutated #1321	GAGTTACAAG CAATTTATCT AGCTTGCAG GATTGGGAT GAGCTGCAAG CCATCTACCT AGCTCTGCAA GACAGCGGAC .....*
>wildtype >mutated #1361	TAGAAGTAAA CATACTAAC AACTCACAAAT ATGCATTAGG TGGAAAGTCAA CATCGTGACA GACTCACAGT ACGCACTGGG .....*
>wildtype >mutated #1401	AATCATTCAA GCACAACCAAG ATCAAAGTGA ATCAGAGTTA CATCATCCAA GCACAACCAAG ACCAATCCGA GTCAGAGCTG .....*
>wildtype >mutated #1441	GTCAATCAA TAATAGAGCA GTTAATAAAA AAGGAAAAGG GTGAACCAGA TCATCGAGCA GCTGATCAAG AAGGAGAAAG .....*
>wildtype >mutated #1481	TCTATCTGGC ATGGGTACCA GCACACAAAG GAATTGGAGG TGTACCTGGC ATGGGTACCA GCACACAAAG GAATTGGAGG .....*
>wildtype >mutated #1521	AAATGAACAA GTAGATAAAT TAGTCAGTGC TGGAAATCAGG AAATGAACAA GTAGATAAAT TAGTCAGTGC TGGGATCCGG .....*
>wildtype >mutated #1561	AAAGTACTAT TTTTAGATGG AATAGATAAG GCCCAAGATG AAGGTGCTGT TCCTGGACGG GATCGATAAG GCCCAAGATG .....*
>wildtype >mutated #1601	AACATGAGAA ATATCACAGT AATTGGAGAG CAATGGCTAG AACATGAGAA GTACCACTCC AACTGGCGCG CTATGGCCAG .....*
>wildtype >mutated #1641	TGATTTAAC CTGCCACCTG TAGTAGCAAA AGAAATAGTA CGACTTCAAC CTGCCACCTG TAGTAGCAAA AGAAATAGTA .....*
>wildtype >mutated #1681	GCCAGCTGTG ATAAATGTCA GCTAAAAGGA GAAGCCATGC GCCAGCTGTG ATAAATGTCA GCTAAAAGGA GAAGCCATGC .....*
>wildtype >mutated #1721	ATGGACAAGT AGACTGTAGT CCAGGAATAT GGCAACTAGA ATGGACAAGT AGACTGTAGT CCAGGAATAT GGCAACTAGA .....*

>wildtype	TTGTACACAT TTAGAAGGAA AAGTTATCCT GGTAGCAGTT
>mutated	CTGCACGCAC CTGGAGGGGA AGGTGATCCT GGTAGCAGTT
#1761	.....
>wildtype	CATGTAGCCA GTGGATATAT AGAACAGAAA GTTATTCCAG
>mutated	CATGTAGCCA GTGGATATAT AGAACAGAAA GTTATCCCTG
#1801	.....
>wildtype	CAGAAACAGG GCAGGAAACA GCATATTTTC TTTTAAAAATT
>mutated	CTGAAAATGG GCAGGAAACA GCATATTTTC TTTTAAAAATT
#1841	.....
>wildtype	AGCAGGAAGA TGGCCAGTAA AAACAATACA TACAGACAAT
>mutated	AGCAGGAAGA TGGCCAGTAA AAACAATACA CACGGACAAC
#1881	.....
>wildtype	GGCAGCAATT TCACCAAGTGC TACGGTTAAG GCCGCCTGTT
>mutated	GGAAGCAACT TCACTGGTGC TACGGTTAAG GCCGCCTGTT
#1921	.....
>wildtype	GGTGGGCGGG AATCAAGCAG GAATTGGAA TTCCCTACAA
>mutated	GGTGGGCGGG AATCAAGCAG GAATTGGAA TTCCCTACAA
#1961	.....
>wildtype	TCCCCAAAGT CAAGGAGTAG TAGAATCTAT GAATAAAGAA
>mutated	TCCCCAATCG CAAGGAGTCG TGGAGAGCAT GAACAAGGAG
#2001	.....
>wildtype	TTAAAGAAAA TTATAGGACA GGTAAGAGAT CAGGCTGAAC
>mutated	CTGAAGAAGA TCATCGGACA AGTGAGGGAT CAGGCTGAGC
#2041	.....
>wildtype	ATCTTAAGAC AGCAGTACAA ATGGCAGTAT TCATCCACAA
>mutated	ACCTGAAGAC AGCAGTGCAG ATGGCAGTGT TCATCCACAA
#2081	.....
>wildtype	TTTTAAAAGA AAAGGGGGGA TTGGGGGGTA CAGTGCAGGG
>mutated	CTTCAAAAGA AAAGGGGGGA TTGGGGGGTA CAGTGCAGGG
#2121	.....
>wildtype	GAAAGAATAG TAGACATAAT AGAACACAGAC ATACAAACTA
>mutated	GAAAGGATCG TGGACATCAT CGCCACCGAC ATCCAAACCA
#2161	.....

10/45

>wildtype >mutated #2201	AAGAATTACA AAAACAAATT ACAAAAATTC AAAATTTCG AGGAGCTGCA GAAGCAGATC ACCAAGATCC AGAACTTCCG .....*
>wildtype >mutated #2241	GGTTTATTAC AGGGACAGCA GAAATCCACT TTGGAAAGGA GGTGTACTAC CGCGACAGCC GCAACCCACT GTGGAAGGGA .....*
>wildtype >mutated #2281	CCAGCAAAGC TCCTCTGGAA AGGTGAAGGG GCAGTAGTAA CCAGCAAAGC TCCTCTGGAA GGGAGAGGGG GCAGTGGTGA .....*
>wildtype >mutated #2321	TACAAGATAA TAGTGACATA AAAGTAGTGC CAAGAAGAAA TCCAGGACAA CAGTGACATC AAAGTGGTGC CAAGGCGCAA .....*
>wildtype >mutated #2361	AGCAAAGATC ATAGGGATT ATGGAAAACA GATGGCAGGT GGCCAAGATC ATCCGCGACT ATGGAAAACA GATGGCAGGT .....*
>wildtype >mutated #2401	GATGATTGTG TGGCAAGTAG ACAGGATGAG GATTAGAAC GATGATTGTG TGGCAAGTAG ACAGGATGAG GATTAGAAC .....*
>wildtype >mutated #2441	TGGAAAAGTT TAGTAAAACA CCATATG TGGAAAGAGCC TGGTGAAGCA CCATATG .....*

ATGGGCGTGAGAAACTCCGTCTTGTCAAGGAAGAAAGCAGATGAATTAG  
AAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGC  
ATGTAGTATGGGCAGCAAATGAATTAGATAGATTGGATTAGCAGAAAG  
CCTGTTGGAGAACAAAGAAGGATGTCAAAAAACTTTCGGCTTAGCT  
CCATTAGTGCCAACAGGCTCAGAAAATTAAAAAGCCTTATAAACTG  
TCTCGTCATCTGGTGCATTACGCAGAACAGAAAGTAAACACACTGA  
GGAAGCAAAACAGATAGTCAGAGACACCTAGTGGTGGAAACAGGAAC  
CACCGAAACCATGCCGAAGACCTCTCGACCAACAGCACCCTAGCGGC  
AGAGGAGGAAACTACCCAGTACAGCAGATCGGTGGCAACTACGTCCAC  
CTGCCACTGTCCCCGAGAACCCCTGAACGCTTGGTCAAGCTGATCGAGG  
AGAAGAAGTCGGAGCAGAACAGTACTGCCAGGATTCCAGGCACTGTCAG  
AAGGTTGCACCCCTACGACATCAACCAGATGCTGAACGCTTGGAGA  
CCATCAGGCGGCTATGCAGATCATCCGTGACATCATCAACGAGGAGGCT  
GCAGATTGGACTTGCAGCACCCACAACCAGCTCCACAACAAGGACAA  
CTTAGGGAGCCGTCAGGATCAGACATCGCAGGAACCACCTCCTCAGTTG  
ACGAACAGATCCAGTGGATGTACCGTCAGCAGAACCCGATCCCAGTAGG  
CAACATCTACCGTCGATGGATCCAGCTGGGCTGCAGAAATGCGTCCGT  
ATGTACAACCCGACCAACATTCTAGATGTAAAACAAGGGCCAAAAGAG  
CCATTTCAGAGCTATGTAGACAGGTTCTACAAAAGTTAAGAGCAGAAC  
AGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACACTGCTGATTCA  
AAATGCTAACCCAGATTGCAAGCTAGTGTCAAGGGGGCTGGGTGTGAAT  
CCACCCCTAGAAGAAATGCTGACGGCTGTCAAGGAGTAGGGGGCCG  
GGACAGAAGGCTAGATTAAATGGCAGAACGCCCTGAAAGAGGCCCTCGCA  
CCAGTGCCAATCCCTTTGCAGCAGCCAACAGAGGGGACCAAGAAAGC  
CAATTAAAGTGTGGAATTGTGGAAAGAGGGACACTCTGCAAGGCAATG  
CAGAGCCCCAACAGAACAGGGATGCTGGAAATGTGGAAAATGGACCA  
TGTATGCCAAATGCCAACAGACAGACAGGGGGTTTTAGGCCTTGGT  
CCATGGGAAAGAAGCCCCGCAATTCCCCATGGCTCAAGTGCATCAGG  
GGCTGATGCCAACTGCTCCCCAGAGGACCCAGCTGTGGATCTGCTAAA  
GAACATGCAGTTGGCAAGCAGCAGAGAGAAAAGCAGAGAGAAAAG  
CAGAGAGAACCTTACAAGGAGGTGACAGAGGATTGCTGCACCTCAAT  
TCTCTTTGGAGGGAGACCACTAG

FIG. 3

SIV gag -----  
 #1 .....  
 ATGGGCGTGAGAAACTCCGTCTTGTCAAGGAAGAAAGCAG

SIV gag -----  
 #41 .....  
 ATGAATTAGAAAAAATTAGGCTACGACCCAACGGAAAGAA

SIV gag -----  
 #81 .....  
 AAAGTACATGTTGAAGCATGTAGTATGGCAGCAAATGAA

SIV gag -----  
 #121 .....  
 TTAGATAGATTGGATTAGCAGAAAGCCTGTTGGAGAACAA

SIV gag -----  
 #161 .....  
 AAGAAGGATGTCAAAAATACTTCGGTCTAGCTCCATT

SIV gag -----  
 #201 .....  
 AGTGCCAACAGGCTCAGAAAATTAAAAAGCCTTATAAT

SIV gag -----  
 #241 .....  
 ACTGTCTGCGTCATCTGGTGCATTACGCAGAAGAGAAAG

SIV gag -----  
 SIVgagDX.. -----  
 #281 .....  
 TGAAACACACTGAGGAAGCAAAACAGATAGTCAGAGACA

SIV gag -----A--A-----T----A--A  
 SIVgagDX..-----C--C-----C----G--G  
 #321 .....  
 CCTAGTGGAAACAGGAACMACMGAAACYATGCCRAAR

SIV gag --AAG-A-----  
 SIVgagDX..---CTC-C-----  
 #361 .....  
 ACMWSTMGACCAACAGCACCATCTAGCGGCAGAGGAGGAA

FIG. 4

SIV gag -T-----A-A-A-----T-----  
 SIVgagDX..-C-----G-G-C-----C-----  
 #401 .....  
 AYTACCCAGTACARCARATMGGTGGTAACTAYGTCCACCT

SIV gag -----T-AAG-----AT-A-T-C-----A--AT--  
 SIVgagDX...---C-GTC-----CC-G-C-T----C--GC--  
 #441 .....  
 GCCAYTRWSCCGAGAACMYTRAAYGCYTGGGTMAARYTG

SIV gag --A-----A-----A--T-----  
 SIVgagDX...-C-----G-----G-C-----  
 #481 .....  
 ATMGAGGARAAGAARTTYGGAGCAGAAGTAGTAGTGCCAGGAT

SIV gag -T-----T-----T-----T-----  
 SIVgagDX..-C-----C-----C-----C-----  
 #521 .....  
 TYCAGGCACTGTCAGAAGGTTGCACCCCCTAYGACATYAA

SIV gag T-----T-A--T--T--G-----A-----  
 SIVgagDX..C-----C-G--C--C--T-----G-----  
 #561 .....  
 YCAGATGYTRAAYTGYGKGGAGACCATCARGCGGCTATG

SIV gag -----T--A-A--T--T--A-----  
 SIVgagDX...---C--C-T--C--C--C-----  
 #601 .....  
 CAGATYATCMGWGAYATYATMAACGAGGAGGCTGCAGATT

SIV gag -----  
 SIVgagDX...-----  
 #641 .....  
 GGGACTTGCAGCACCCACAACCAAGCTCCACAACAAGGACA

SIV gag -----T--T-----A--T-----  
 SIVgagDX...-----C--C-----C--C-----  
 #681 .....  
 ACTTACGGAGCCGTCAAGGATCAGAYATYGCAGGAACMACY

SIV gag AGT-----A--T-----A-----A-A--A--  
 SIVgagDX..TCC-----T--C-----G-----C-T--G--  
 #721 .....  
 WSYTCAGTWGAYAACARATCCAGTGGATGTACMGWCARC

SIV gag -----C-A-----T-A-GA-----  
 SIVgagDX.....G-C-----C-C-TC-----  
 #761 .....  
 AGAACCCSATMCCAGTAGGCAACATYACMGKMGATGGAT

SIV gag ---A----GT---A--A--T--CA-A----T----A  
 SIVgagDX...G---TC---G--G--C--TC-T---C---G  
 #801 .....  
 CCARCTGGGKYTGCARAARTGYGTYMGWATGTAYAACCR

SIV gag --A-----  
 SIVgagDX...C-----  
 #841 .....  
 ACMAACATTCTAGATGTAAAACAAGGGCCAAAGAGCCAT

SIV gag -----  
 #881 .....  
 TTCAGAGCTATGTAGACAGGTTCTACAAAAGTTAACAGC

SIV gag -----  
 #921 .....  
 AGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAA

SIV gag -----  
 #961 .....  
 ACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAG

SIV gag -----  
 #1001 .....  
 TGCTGAAGGGCTGGGTGTGAATCCCACCCCTAGAAGAAAT

SIV gag -----  
 #1041 .....  
 GCTGACGGCTTGTCAAGGAGTAGGGGGCCGGACAGAAG

SIV gag -----  
 #1081 .....  
 GCTAGATTAATGGCAGAAGCCCTGAAAGAGGCCCTCGCAC

SIV gag -----  
 #1121 .....  
 CAGTGCCAATCCCTTTGCAGCAGCCAACAGAGGGACC

SIV gag -----  
 #1161 .....  
 AAGAAAGCCAATTAAGTGTGGAATTGTGGAAAGAGGGGA

SIV gag #1201	----- ..... CACTCTGCAAGGCAATGCAGAGCCCCAAGAAGACAGGGAT
SIV gag #1241	----- ..... GCTGGAAATGTGGAAAAATGGACCATGTTATGGCCAAATG
SIV gag #1281	----- ..... CCCAGACAGACAGGCGGGTTTTAGGCCTTGGTCCATGG
SIV gag #1321	----- ..... GGAAAGAAGCCCCGCAATTCCCCATGGCTCAAGTGCATC
SIV gag #1361	----- ..... AGGGGCTGATGCCAACTGCTCCCCAGAGGACCCAGCTGT
SIV gag #1401	----- ..... GGATCTGCTAAAGAACTACATGCAGTTGGCAAGCAGCAG
SIV gag #1441	----- ..... AGAGAAAAGCAGAGAGAAAGCAGAGAGAACCTTACAAGG
SIV gag #1481	----- ..... AGGTGACAGAGGGATTGCTGCACCTCAATTCTCTTTGG
SIV gag #1521	----- ..... AGGAGACCAGTAG

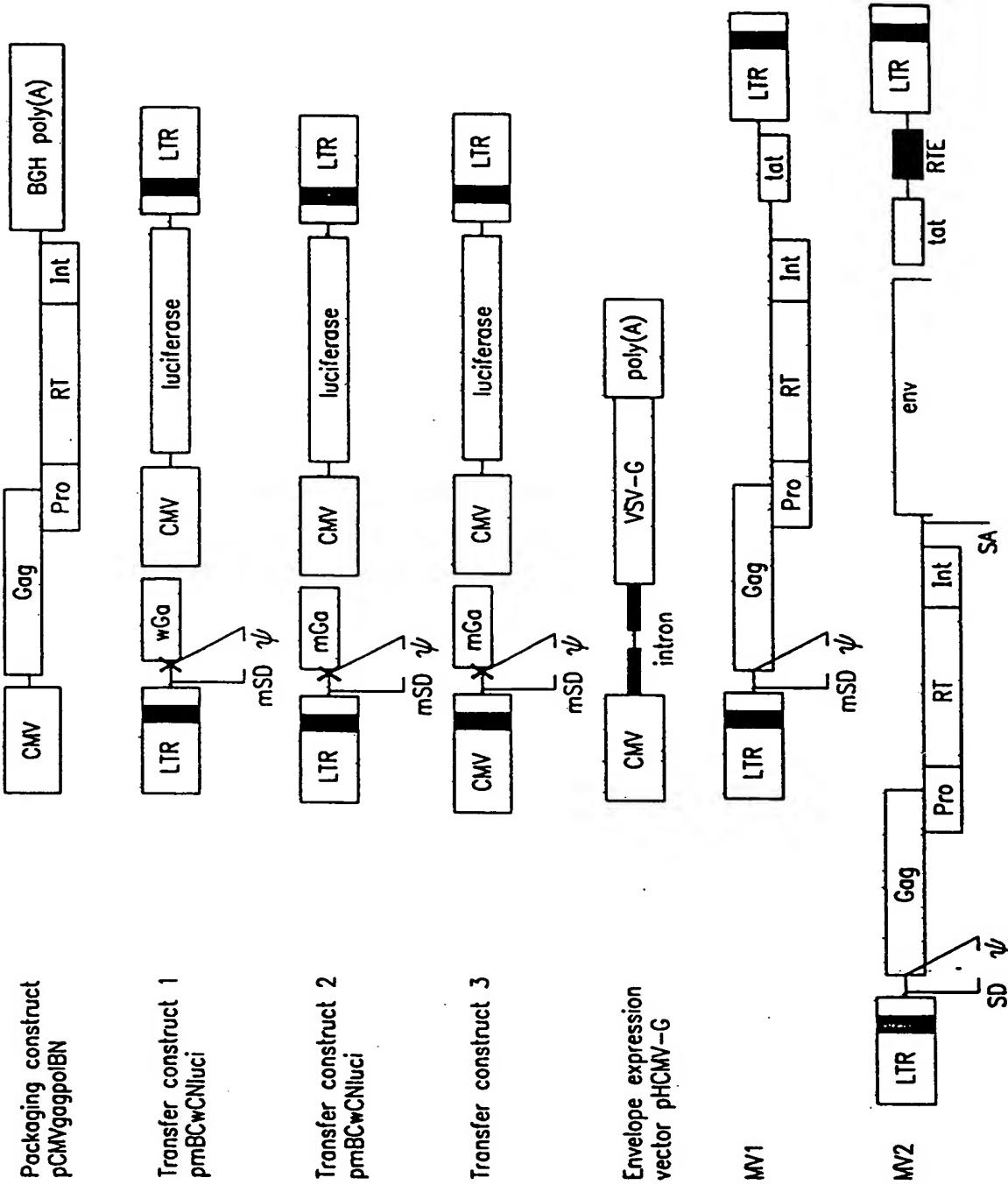


FIG. 5

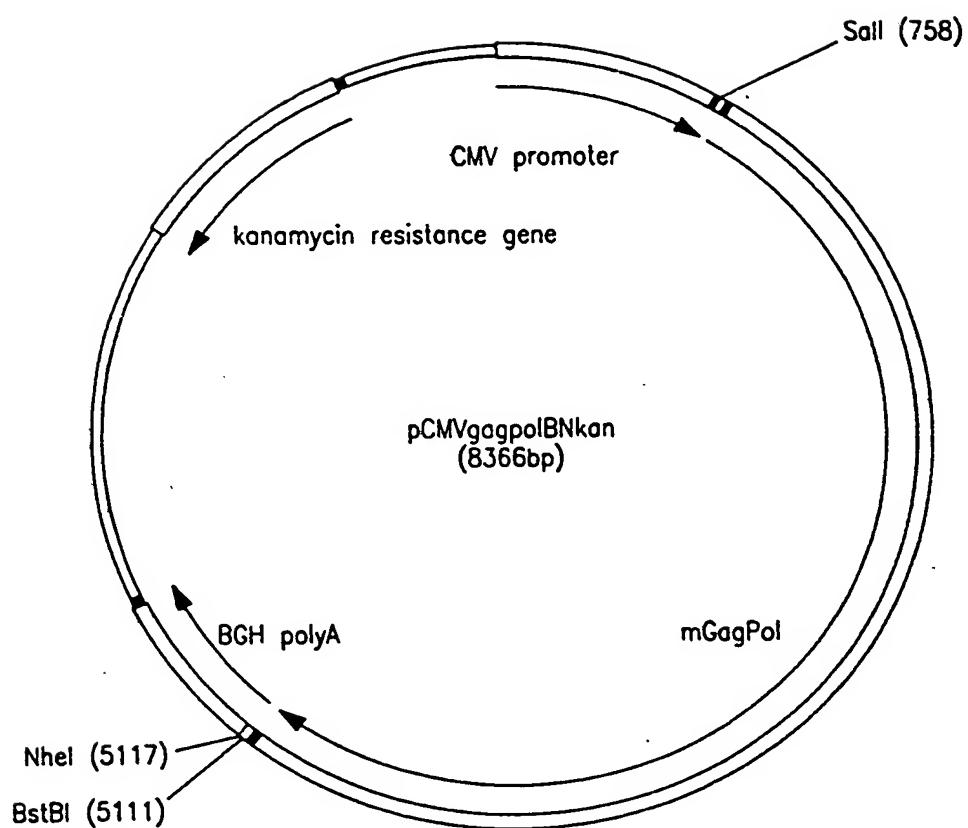


FIG. 6

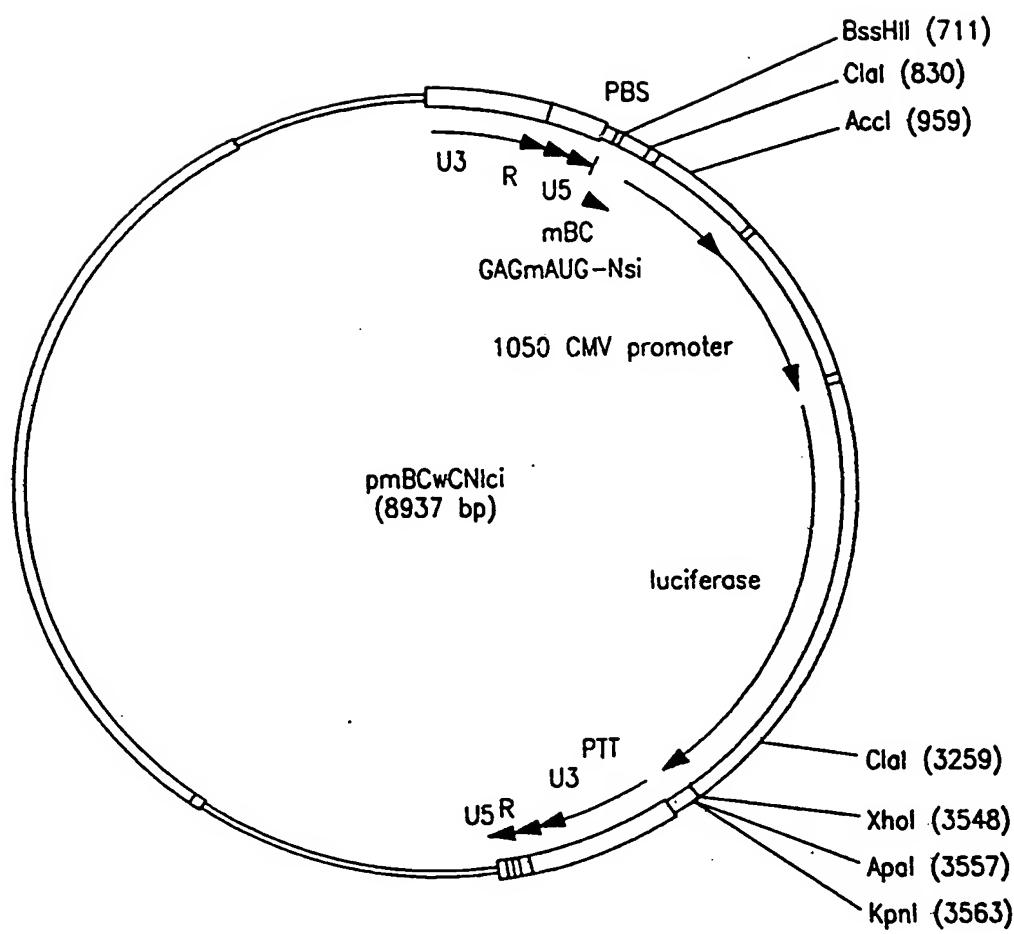


FIG. 7

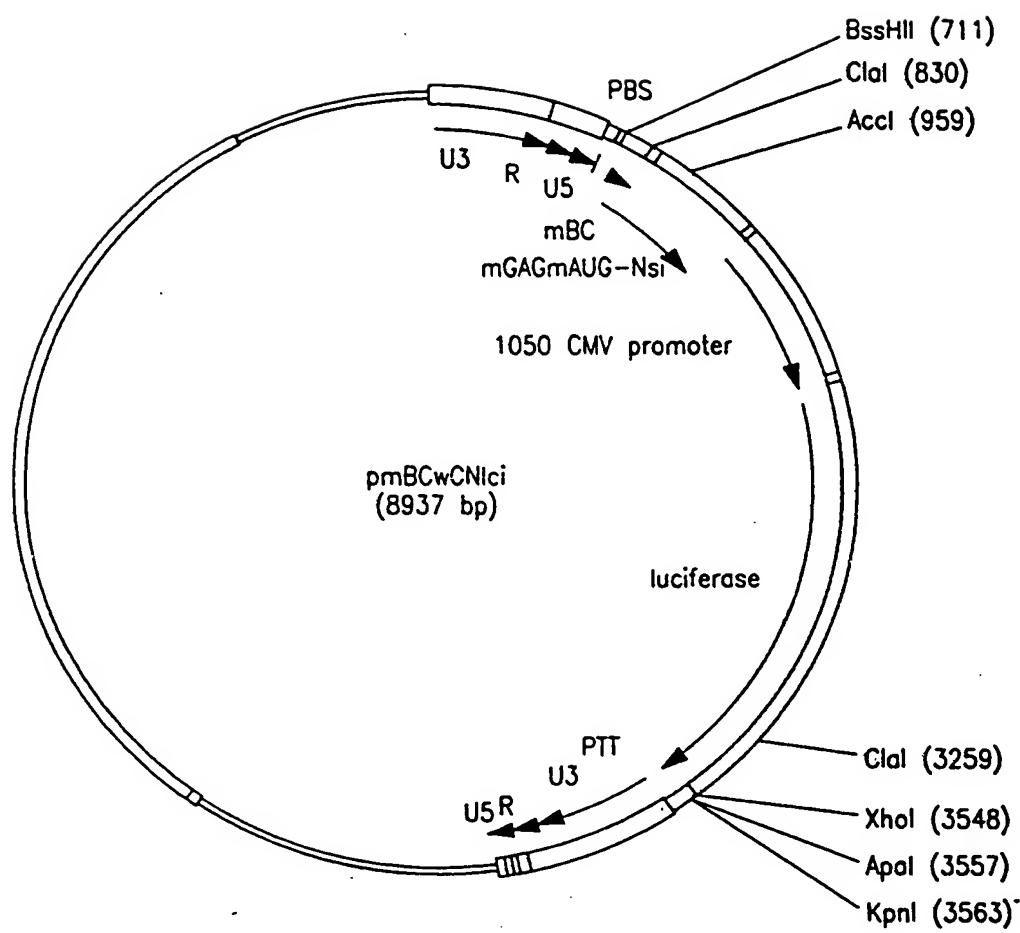


FIG. 8

1 CCTGGCCATT GCATACGTG TATCCATATC ATAATATGTA CATTATATT GGCTCATGTC CAACATTACC  
 71 GCCATGTTGA CATTGATTAT TGACTAGTTA TTAATAGTAA TCAATTACGG CGTCATTAGT TCATAGCCCA  
 141 TATATGGACT TCCCGGTAC ATAACCTACG GAAATGGCC CGCCTGGCTG ACCGCCAAC GACCCCCGCC  
 211 CATTGACGTC AATAATGAGC TATGTTCCA TAGAACGCC AATAGGGACT TTCCATTGAC GTCAATGGGT  
 281 GGACTATTAA CGGTAAACTG CCCACTTGGC ACTACATCAA GTGTATCATA TGCCAACTAC GCCCCCTATT  
 351 GACGTCAATG ACGGTAAATG GCCCCCTGG CATTATGCC AGTACATGAC CTATGGAC TTCTACTT  
 421 GCCAGTACAT CTACGTATTA GTCATCGCTA TTACCATGGT GATGCGGTTT TGGCAGTACA TCAATGGCG  
 491 TGGATAGCGG TTGACTCAC GGGGATTCC AAGTCTCCAC CCCATTGACG TCAATGGAG TTGTTTGG  
 561 CACCAAAATC AACCGGACTT TCCAAAATGT CGTAACAACT CCCCCCCATT GACGAAATG GGCGTAGGC  
 631 GTGTACGGTG GGAGGTCTAT ATAAGCAGAG CTCGTTAGT GAACCGTCAG ATCGCCTGGA GACGCCATCC

SalI (758)

701 ACGCTGTTT GACCTCCATA GAAGACACCG GGACCGATCC AGCCTCCGGC GGCGCGCGTC GACAGAGAGA →  
 771 TGGGTGGAG AGCGTCAGTA TTAAGGGGG GAGAATTAGA TCGATGGAA AAAATTGCGT TAAGGCCAGG  
 841 GGGAAACAAG AAGTACAAGC TAAACACAT CGTATGGCA ACCAGGGAGC TAGAACGATT CGCAGTTAAT  
 911 CCTGGCCGTG TAGAACATC AGAAGGCTGT AGACAAATAC TGGGACAGCT ACAACCATCC CTTCAGACAG  
 981 GATCAGAGGA GCTTCGATCA CTATACAACA CAGTAGCAAC CCTCTATTGT GTGCACCAGC GGATCGAGAT  
 1051 CAAGGACACC AAGGAAGCTT TAGACAAGAT AGAGGAAGAG CAAACAAAGT CCAAGAAGAA GGCCCAGCAG  
 1121 CGACGAGCTG ACACAGGACA CAGCAATCAG GTCAGCCAAA ATTACCCCTAT AGTGCAGAAC ATCCAGGGCC  
 1191 AAATGGTACA TCAGGCCATA TCACCTAGAA CTTTAAATGC ATGGTAAAAA GTAGTAGAAG AGAAGGCTT  
 1261 CAGCCCAGAA GTGATACCCA TGTTTCAGC ATTATCAGAA GGAGCCACCC CACAGGACCT GAACACGATG  
 1331 TTGAACACCG TGGGGGACA TCAAGCAGCC ATGAAATGT TAAAGAGAC CATCAATGAG GAAGCTGCAG  
 1401 AATGGGATAG AGTCATCCA GTGCATCCAG GGCTATTCC ACCAGGCCAG ATGAGAGAAC CAAGGGGAAG  
 1471 TGACATAGCA GGAACACTA GTACCCCTCA GGAACAAATA GGATGGATGA CAAATAATCC ACCTATCCCA  
 1541 GTAGGAGAGA TCTACAAGAG GTGGATAATC CTGGGATTGA ACAAGATCGT GAGGATGTAT AGCCCTACCA  
 1611 GCATTCTGGA CATAAGACAA GGACCAAAGG AACCTTTAG AGACTATGTA GACCGTTCT ATAAAACCT  
 1681 AAGAGCTGAG CAAGCTTCAC AGCAGGTAAA AAATGGATG ACAGAAACCT TGTGGTCCA AAATGCCAAC  
 1751 CCAGATTGTA AGACCATCCT GAAGGCTCTC GCCCCAGCGG CTACACTAGA AGAAATGATG ACAGCATGTC  
 1821 AGGGAGTAGG AGGACCCGGC CATAAGGCAA GAGTTTGGC CGAGGGATG AGCCAGGTGA CGAACTCGGC

FIG. 9A

1891	GACCATAATG ATCCAGAGAG GCAACTTCCG GAACCAGCGG AAGATCGTCA ACTGCCTCAA TTGTGGCAA
1961	GAAGGGCACA CGGCCAGGA CTGGCGGCC CCCCGGAAGA AGGGCTGTTG GAAATGTGGA AAGGAAGGAC
2031	ACCAAATGAA AGATTGTACT GAGAGACAGG CTAATTTTT AGGGAAGATC TGGCCTTCCT ACAAGGGAAG
2101	GCCAGGGAAT TTTCTTCAGA CCAGACCAGA GCCAACAGCC CCACCAGAAC AGAGCTTCAG GTCTGGGTA
2171	GAGACAACAA CTCCCCCTCA GAAGCAGGAG CCGATAGACA AGGAAGCTGA TCCTTTAACT TCCCTCAGAT
2241	CACTCTTGG CAACGACCCC TCGTCACAGT AAGGATCGGG GGGCAACTCA AGGAAGCGCT GCTCGATA
2311	GGAGCAGATG ATACAGTATT AGAAGAAATG AGTTGCCAG GAAGATGGAA ACCAAAAATG ATAGGGGGA
2381	TCGGGGCTT CATCAAGGTG AGGCAGTACG ACCAGATACT CATAGAAATC TGTGGACATA AAGCTATAGG
2451	TACAGTATTA GTAGGACCTA CACCTGTCAA CATAATTGGA AGAAATCTGT TGACCCAGAT CGGCTGCACC
2521	TTGAACCTCC CCATCAGCCC TATTGAGACG GTGCCCCGTGA AGTTGAAGCC GGGGATGGAC GGCCCCAAGG
2591	TCAAGCAATG GCCATTGACG AAAGAGAAGA TCAAGGCCTT AGTCGAAATC TGTACAGAGA TGGAGAAGGA
2661	AGGGAAAGATC ACCAAGATCG CCCCTGAGAA CCCCTACAAAC ACTCCAGTCT TCGCAATCAA GAAGAAGGAC
2731	AGTACCAAGT GGAGAAAGCT GGTGGACTTC AGAGAGCTGA ACAAGAGAAC TCAGGACTTC TGGGAAGTTC
2801	AGCTGGGCAT CCCACATCCC GCTGGGTGGA AGAAGAAGAA GTCAAGTGACA GTGCTGGATG TGGGTGATGC
2871	CTACTTCTCC CTTCCCTTGG ACCAGGACTT CAGGAAGTAC ACTGCCTTCAG CGATACCTAG CATCAACAAAC
2941	GAGACACCAG GCATCCGCTA CCAGTACAAC GTGCTGCCAC AGGGATGGAA GGGATCACCA GCCATCTTC
3011	AAAGCAGCAT GACCAAGATC CTGGACCCCT TCCCGAAGCA AAACCCAGAC ATCGTGTCT ATCACTACAT
3081	GGACGACCTC TACGTAGGAA GTGACCTGGA GATCGGGCAG CACAGGACCA AGATCGAGGA GCTGAGACAG
3151	CATCTGTTGA GGTGGGACT GACCACACCA GACAAGAAC ACCACAAGGA ACCTCCCTTC CTGTGGATGG
3221	GCTACCAACT GCATCCTGAC AAGTGGACAG TGCAAGCCAT CGTGTGCT GAGAAGGACA GCTGGACTGT
3291	GAACCGACATA CAGAAGCTCG TGGGCAAGTT GAACTGGCA AGCCAGATCT ACCCAGGCAT CAAAGTTAGG
3361	CAGCTGTGCA AGCTGCTTCG AGGAACCAAG GCACTGACAG AAGTGTCTCC ACTGACAGAG GAACCGAGGC
3431	TAGAACTGGC AGAGAACCGA GAGATCCTGA AGGAGCCAGT ACATGGAGTG TACTACGACC CAACCAAGGA
3501	CCTGATCGCA GAGATCCAGA AGCAGGGCA AGGCCAATGG ACCTACAAA TCTACCAGGA GCCCTTCAAG
3571	AACCTGAAGA CAGGCAAGTA CGCAAGGATG AGGGGTGCC ACACCAACGA TGTGAAGCAG CTGACAGAGG
3641	CAGTGCAGAA GATCACCACCA GAGACCACCG TGATCTGGGG CAAGACTCCC AAGTTCAAGC TGCCCATACA
3711	GAAGGGAGACA TGGGAGACAT GGTGGACCGA GTACTGGCAA GCCACCTGGA TCCCTGAGTG CGAGTTCGTG

FIG. 9B

3781 AACACCCCTC CCTTGGTGA ACTGTGGTAT CAGCTGGAGA AGGAACCCAT CGTGGGAGCA GAGACCTTCT  
 3851 ACCTGGATGG GGCAGCCAAC AGGGAGACCA AGCTGGCAA GGCAGGCTAC GTGACCAACC GAGGACCGACA  
 3921 CAAAGTGGTG ACCCTGACTG ACACCACCAA CCAGAACACT GAGCTGCAAG CCATCTACCT AGCTCTGCAA  
 3991 GACAGCCGAC TGGAACTGAA CATCGTGACA CACTCACAGT ACCCACTGGG CATCATCCAA GCACAACCCAG  
 4061 ACCAATCCCA GTCAGAGCTG GTGAACCAGA TCATCGAGCA GCTGATCAAG AAGGAGAAAG TGTACCTGGC  
 4131 ATGGGTACCA GCACACAAAG GAATTGGAGG AAATGAACAA GTAGATAAT TAGTCAGTGC TGGGATCCGG  
 4201 AAGGTGGTGT TCCTGGACGG GATCGATAAG GCCCAAGATG AACATGAGAA GTACCACTCC AACTGGCCGG  
 4271 CTATGCCAG CGACTTCAAC CTGCCACCTG TAGTAGCAA AGAAATAGTA GCCAGCTGTG ATAAATGTCA  
 4341 GCTAAAAGGA CAAGCCATGC ATGGACAAGT AGACTGTAGT CCAGGAATAT GGCAACCTGGA CTGCACGCAC  
 4411 CTGGAGGGGA AGGTGATCCT GGTAGCACTT CATGTAGCCA GTGGATATAT AGAACAGAA GTTATCCCTG  
 4481 CTGAAACTGG GCAGGAAACA GCATATTTTC TTTAAAATT ACCAGGAACA TGCCCACTAA AAACAATACA  
 4551 CACCGACAAC GGAAGCAACT TCACTGGTGC TACGGTTAAG GCGCCCTGTT GGTGGGCGGG AATCAACCG  
 4621 CAATTTGAA TTCCCTACAA TCCCCAATCG CAAGGAGTCG TGGAGAGCAT GAACAAGGAG CTGAAGAAGA  
 4691 TCATCGGACA AGTGAGGGAT CAGGCTGAGC ACCTGAAGAC ACCAGTGCAG ATGGCAGTGT TCATCCACAA  
 4761 CTTCAAAAGA AAACGGGGGA TTGGGGGTA CAGTGCAGGG GAAAGGATCG TGGACATCAT CGCCACCCGAC  
 4831 ATCCAAACCA ACCGAGCTGCA GAAGCAGATC ACCAAGATCC AGAACTTCCG GGTGTACTAC CGGGACACCC  
 4901 GCAACCCACT GTCCAAGGGGA CCAGCAAAGC TCCTCTGGAA CGGAGAGGGG GCAGTGGTGA TCCAGGACAA  
 4971 CAGTCACATC AAAGTGGTGC CAAGGCCAA GGCAAGATC ATCCGGACT ATGAAAACA GATGGCAGGT  
 5041 GATGATTGTG TGGCAAGTAG ACAGGATGAG GATTAGAACCC TGGAAAGAGCC TGGTGAAGCA CCATATGGCC

NheI (5117) →  
BstBI (5111)

5111 TTCCGAAGCTA CCCTCGAGAT CCAGATCTGC TGTGCCCTCT AGTTGCCAGC CATCTGTGT TGGCCCTCC  
 5181 CCCGTGCCCT CCTTGACCCCT GGAACGGGCC ACTCCCCTG TCCTTCTCTA ATAAAATGAG GAAATTGCAT  
 5251 CGCATTGTCT CACTAGGTGT CATTCTATTG TGCGGGGTGG GGTGGGCGAG CACAGCAAGG GGGAGGATTG  
 5321 GGAAGACAAT AGCAGGCATG CTGGGGATGC GGTGGGCTCT ATGCCTACCC AGGTGCTGAA GAATTGACCC  
 5391 GGTTCCCTCT GGGCCAGAAA GAAGCAGGCA CATCCCCTTC TCTCTGACAC ACCCTGTCCA CGCCCTGGT  
 5461 TCTTACTTCC AGCCCCACTC ATAGGACACT CATACTGAG GAGGGCTCCG CCTTCAATCC CACCCGCTAA  
 5531 AGTACTTCCA CGGGTCTCTC CCTCCCTCAT CAGCCCCACCA AACCAAACCT AGCCTCCAAG AGTGGGAAGA

5601 AATTAAGCA AGATAGGCTA TTAAGTGCAG AGGGAGAGAA AATCCCTCCA ACATGTGAGG AAGTAATGAG  
 5671 AGAAAATCATA GAATTTCTTC CCGCTTCCCG CTCACTGACT CGCTGCCCTC GGTGGTTCGG CTGGCGGAG  
 5741 CCGGTATCAGC TCACTCAAAG GCGCTAATAC CGGTATCCAC AGAACAGGG CATAACCGAG CAAAGAACAT  
 5811 GTGAGCAAAA GGGCAGCAAA AGCCCAGGAA CGCTAAAAAG GCCCCGTTGC TGCGGTTTTT CCATAGGCTC  
 5881 CGCCCCCTG ACCAGCATCA CAAAAATCGA CGCTCAAGTC AGAGGTGGCG AAACCCGACA GGACTATAAA  
 5951 GATACCAGGC GTTCTCCCTC GCAAGCTCCC TCGTGCCTCTC TCCTGTTCCG ACCCTGCCG TTACCGGATA  
 6021 CCTGTCGGCC TTCTCCCTT CGGGGAAGCGT GGGGCTTCT CAATGCTCAC GCTGTAGGTA TCTCAGTTCG  
 6091 GTGTAGGTGG TTCGCTCCAA GCTGGCTGT CTGGCAGAAC CCCCCGTTCA GCGGACCCG TGCGCCATTAT  
 6161 CCCGTAACTA TCGTCTTGAG TCCAACCCGG TAAGACACGA CTTATGCCCA CTGGCAGCAG CCACCTGCTA  
 6231 CAGGATTAGC AGAGCCAGGT ATGTAGGCGG TCCTACAGAG TTCTGAAGT GGTGGCTAA CTACGGCTAC  
 6301 ACTAGAAGGA CAGTATTG TATCTCGCT CTGCTGAAGC CAGTTACCTT CGGAAAAAGA GTGGTAGCT  
 6371 CTTGATCCCG CAAACAAAC ACCGCTGCTA GCGGTGTTT TTTGTTGCT AAGCAGCAGA TTACGGCCAG  
 6441 AAAAAAAGGA TCTCAAGAAG ATCCTTGTAT CTTTCTACCG GGGCTGAGC CTCAGTGGAA CGAAAAGTC  
 6511 CGTTAAGGG A TTTGGTCAT CAGATTATCA AAAAGGATCT TCACCTACAT CTTTTAAAT TAAAAATGAA  
 6581 GTTTAAATC ATATCAAAGT ATATATGAGT AAACCTGCTC TGACACTTAC CAATGCTAA TCAGTGAGGC  
 6651 ACCTATCTCA CGCATCTGC TATTCGTT ATCCATAGCTT GCCTGACTCC GGGGGGGGGG GGCGCTGAGG  
 6721 TCTGCTCGT GAAGAAGGTG TTGCTGACTC ATACAGGCC TGAATGCCCAT CATCATCCAG CCAGAAAGTG  
 6791 AGGGAGGACAC GTTGTATGAG AGCTTGTG TAGGTGGACC AGTTGGTAT TTTGAACTTT RGCTTTGCCA  
 6861 CGGAACGGTC TCGTTGTCG CGAACATCCG TGATCTGATC CTTCAACTCA CGAAAAGTTC GATTATTCA  
 6931 ACAAAAGCCGC CTCCTCCGTA AGTCACCGTA ATGCTCTGCC AGTGTACAA CCAATTAAAC AATTCTGATT  
 7001 AGAAAAACTC ATCGAGCATC AAATGAAACT GCAATTATT CATATCAGGA TTATCAATAC CATATTITG  
 271 ▲ PhePheGlu AspLeuMetL euHisPheGlu nLeuLysAsn MetAspProA snAspIleGly TyrLysGln  
 7071 AAAAGCCGT TTCTGTAATG AAGGAGAAAA CTCACCGAGG CAGTCCATA GGATGCCAAG ATCCTGGTAT  
 248 ▲ PheLeuArgL ysGlnLeuSe rProSerPhe GluGlyLeuC ysAsnTrpLe uileAlaLeu AspGlnTyra  
 7141 CGGCTCTGCA TTCCGACTCG TCCAAACATCA ATACAACCTA TAAATTCCC CTCGTCAAA ATAAGGTTAT  
 224 ▲ rgAspAlaL eGlyValArg GlyValAspI leCysGlyI IeLeuLysGly GluAspPheI leLeuAsnAs  
 7211 CAAGTCAGAA ATCACCATGA GTGACGACTG AATCCGGTGA GAATGGCAA AGCTTATGCA TTTCTTCCA  
 201 ▲ pLeuSerPhe AspGlyHist hrValValSe rAspProSer PheProLeuL euLysHisMe tGluLysTrp  
 7281 GACTTGTCA ACAGGCCAGC CATTACGCTC GTCATCAAAA TCACTCCCAT CAACCAAACC GTTATTCTATT  
 178 ▲ ValGlnGluV alProTrpGly yAsnArgGlu AspAspPheA spSerAlaAs pValLeuGly AsnAsnMet  
 7351 CGTGATTGCG CCTGAGCCGAC AGCAAATACG CGATCGCTGT TAAAGGACA ATTACAAACA GCAATCGAAT  
 154 ▲ rgSerGlnAl aGlnAlaLeu ArgPheValA rgAspSerAs nPheProCys AsnCysValP roIleSerH  
 7421 GCAACCGGGCG CAGGAACACT GCCAGCCAT CAACAATATT TTCACCTGAA TCAGGATATT CTTCTAAATAC  
 131 ▲ sLeuArgArg LeuPheVala laLeuAlaAs pValIleAsn GluGlySerA spProTyrGlu uGluLeuVal  
 7491 CTGGAATGCT GTTTCCCGG CGATCGCAGT GGTGAGTAAAC CATGCCATCAT CAGGAGTACG GATAAAATG  
 108 ▲ GlnPheAlaT hrLysGlyPr oIleAlaThr ThrLeuLeuT rpAlaAspAs pProThrArg IlePheH  
 7561 TTGATGGTCG GAAGAGGCAT AAATTCCGTC AGCCAGTTA GTCTGACCAT CTCATCTGTA ACATCAATTG  
 84 ▲ ysIleThrPr oLeuProMet PheGluThrL euTrpAsnLe uArgValMet GluAspThrV alAspAsnAI  
 7631 CAACGCTACC TTGCCATGT TTCAAGAACAA ACTCTGGCGC ATCGGGCTTC CCATACAATC GATACATTG  
 61 ▲ aValSerGly LysGlyHist ysLeuPheLe uGluProAla AspProLysG lyTyrLeuAr gTyrIleThr  
 7701 CGCACCTGAT TGCCCGACAT TATCGCGAGC CCATTATAC CCATATAAAAT CAGCATCCAT GTGGAATT  
 38 ▲ AlaGlySerG lnGlyValAs nAspArgAla TrpLysTyrG lyTyrLeuAs pAlaAspMet AsnSerAsnL  
 7771 AATCCGGCC TCGAGCAAGA CGTTTCCGT TGAATATGCC TCATACACC CCTTGTATTA CTGTTATG  
 14 ▲ euArgProAr gSerCysSer ThrGluArgG lnIleHisSe rMet  
 7841 AAGCAGACAG TTTTATCTT CATCATGATA TATTTTATC TTGTGCAATG TAACATCAGA GATTTGAGA  
 7911 CACAACTGG CTTTCCCCC CCCCCCTTA TTGAAGCATT TATCAGGGTT ATTGTCTCAT GAGCGGATAC  
 7981 ATATTTGAAT GTATTTAGAA AAATAAACAA ATAGGGCTTC CGGGCACATT TCCCCGAAAAA GTGCCACCTG  
 8051 ACCTCTAACAG AACCAATTATT ATCATGACAT TAAACCTATAA AAATAGGGCT ATCACCGAGGC CTTTCTGCT  
 8121 CGGGCTTC GGTGATGACG GTGAAAACCT CTGACACATG CAGCTCCCCG AGACGGTCAC AGCTTGTCTG  
 8191 TAAGGGATG CCGGGACCAAG ACAACCCGT CAGGGCCGCT CAGGGGGTGT TGGGGGGTGT CGGGGCTGCC  
 8261 TTAACATATGC GGCATCAGAG CAGATTGTAC GTAGACTGCA CCATATGCCG TGTGAAATAC CCCACAGATG  
 8331 CGTAAGGACA AAATACCGCA TCAGATTGCC TATTGG (SEQUENCE ID NO: 6)

1 TGGAAAGGGCT AATTTGGTCC CAAAAAAGAC AACAGATCCT TGATCTGTGG ATCTACCACA CACAAGGCTA  
 71 CTTCCCTGAT TCGGCAGAACT ACACACCAGG GCCAGGGATC AGATATCCAC TGACCTTTGG ATGGTGCTTC  
 141 AAGTTAGTAC CAGTTGAACC AGACCAAGTA GAACACGCCA AATAAGGAGA CAAGAACAGC TTGTTACACC  
 211 CTATGACCCA CCATGGGATG CAGGACCCCG ACGGAGAACT ATTAGTGTGG AAGTTGACA GCCTCCTAGC  
 281 ATTTCGTCAC ATGGCCCCGAG ACCTGCATCC CGACTACTAC AAACACTGCT CACATCGACC TTTCTACAAAC  
 351 GGACTTTCCG CTGGGGACTT TCCAGGGAGG TGTGGCTGG GCGGGACTGG GGAGTGGCGA GCCCTCAGAT  
 421 GCTACATATA ACCAGCTGCT TTTTGCCTCT ACTGGGTCTC TCTGGTTAGA CCAGATCTGA GCCTGGGAGC  
 491 TCTCTGGCTA ACTAGGGAAC CCACTGCTTA AGCCTCAATA AAGCTTGCCT TGACTGCTCA AAGTAGTGTG  
 561 TGCCCCGTCTG TTGTGTGACT CTGGTAACTA CAGATCCCTC AGACCCTTT AGTCAGTGTG GAAAATCTCT  
 631 AGCAGTGGCG CCCGAACAGG GACTTGAAAG CGAAAGTAAA GCCAGAGGAG ATCTCTCGAC GCAGGACTCG  
 701 GCTTGCTGAA CGCCGCacgg caagaggcga ggggcggcgC ctgACgagGa cgccaaaaat tttgactagc

BssHII (711)

771 ggaggctaga aggagagagC TCGGTGGAG ACCGTCAGTA TCAAGCGGGG GAGAATTAGA TCGATGGAA

Clai (830)

841 AAAATTGGT TAAGGCCAGG CGGAAAGAAA AAATATAAAT TAAAACATAT AGTATGGCA ACCAGGGAGC

AccI (959)

911 TAGAACGATT CGCAGTTAAT CCTGGCCTGT TAGAACATC AGAAGGCTGT AGACAAATAC TGGGACAGCT

981 ACAACCATCC CTTCAGACAG GATCAGAAGA ACTTAGATCA TTATATAATA CAGTAGAAC CCTCTATTGT

1051 GTGCATCAAA GGATACAGAT AAAAGACACC AAGGAACCTT TAGACAAGAT AGAGGAAGAG CAAAACAAAA

FIG. 10A

1121 GTAAGAAAAA AGCACAGCAA GCAGCAGCTG ACACAGGACA CAGCAATCAG GTCAGCCAAA ATTACCCSTAT  
 1191 AGTGCAGAAC ATCCAGGGGC AAATGGTACA TCAGGCCATA TCACCTAGAA CTTTAAACGA TAAGCTTGGG  
 1261 AGTTCCGGT TACATAACTT ACGGTAAATG GCCCGCCTGG CTGACCGCCC AACGACCCCC GCCCATTGAC  
 1331 GTCAATAATG ACGTATGTTG CCATACTAACG GCCAATAGGG ACTTTCCATT GACGTCAATG GGTGGAGTAT  
 1401 TTACGGTAAA CTGCCCACTT GGCACTACAT CAAGTGTATC ATATGCCAAG TACCCCCCT ATTGACGTCA  
 1471 ATGACGGTAA ATGCCCGGCC TGGCATTATG CCCAGTACAT GACCTTATGG GACTTTCCTA CTTGGCAGTA  
 1541 CATCTACGTA TTAGTCATCG CTATTACCAT GGTGATGGGG TTTTGGCAGT ACATCAATGG GCGTGGATAG  
 1611 CGGTTTGACT CACGGGGATT TCCAAGTCTC CACCCCATTG ACGTCAATGG GAGTTTGTGTT TGGCACCAAA  
 1681 ATCAACGGGA CTTTCCAAAAA TGTCGTAACA ACTCCGCCCC ATTGACGCCA ATGGCCGGTA GCGGTGTACG  
 1751 GTGGGACGTC TATATAACCA GAGCTCGTTT AGTGAACCGT CAGATCGCCT GGAGACGCCA TCCACGGCTGT  
 1821 TTTGACCTCC ATAGAAGACA CCGACTCTAG AG<sup>Gatcc</sup>ATC TAAGTAAGCT TGGCATTCCG CTACTGTGTT  
 1891 TAAAATGGAA GACGCCAAAAA ACATAAAGAA AGGGCCGGCG CCATTCTATC CTCTAGAGGA TGGAACCGCT  
 1961 GGAGAGCAAC TGCATAAGGC TATGAACAGA TACGCCCTGG TTCCCTGGAAC AATTGCTTTT ACAGATGCCAC  
 2031 ATATCGAGGT GAACATCACG TACGCGGAAT ACTTCGAAT GTCCGTTGGG TTGGCAGAAG CTATGAAACG  
 2101 ATATGGCTG AATACAAATC ACAGAATCGT CGTATGCCAGT GAAAATCTC TTCAATTCTT TATGCCGGTG  
 2171 TTGGGCCGT TATTTATCGG AGTTGCAGTT GCGCCCGCGA ACGACATTAA TAATGAACGT GAATTGCTCA  
 2241 ACAGTATGAA CATTTCGGAG CCTACCGTAG TGTTTGTTC CAAAAAGGGG TTGCAAAAAA TTTTGAACGT  
 2311 GCAAAAAAAA TTACCAATAA TCCAGAAAAT TATTATCATG GATTCTAAAA CGGATTACCA CGGATTTCAG

2381 TCGATGTACA CGTCGTCAC ATCTCATCTA CCTCCGGTT TTAATGAATA CGATTTGTA CCAGAGTCCT  
 2451 TTGATCGTGA CAAAACAATT GCACTGATAA TGAATTCTC TGGATCTACT GGGTACCTA AGGGTGTGGC  
 2521 CCTTCCCAT AGAACTCCCT GCGTCAGATT CTCGCATGCC AGAGATCTA TTTTGGCAA TCAAATCATT  
 2591 CCGGATACTG CGATTTAACG TGTTGTTCCA TTCCATCAGG GTTTGGAAT GTTACTACA CTCGGATATT  
 2661 TGATATGTGG ATTTCGAGTC GTCTTAATGT ATAGATTGA AGAAGAGCTG TTTTACGAT CCCTTCAGGA  
 2731 TTACAAAATT CAAAGTGCCT TGCTAGTACC ACCCTATT TCATTCTCG CCAAAAGCAC TCTGATTGAC  
 2801 AAATACGATT TATCTAATT ACACGAAATT GCTTCTGGGG GCGCACCTCT TTGAAAGAA GTCGGGGAAAG  
 2871 CGGTTGAAA ACGCTTCCAT CTTCCAGGGA TACGACAAGG ATATGGGCTC ACTGAGACTA CATCAGCTAT  
 2941 TCTGATTACA CCCGAGGGGG ATGATAAACCG GGGCGCGTC GGTAAAGTTG TTCCATTTC TGAAGCGAAC  
 3011 CTIGTGGATC TGGATACCGG GAAAACGCTG GGCGTTAAC AGAGAGGGCA ATTATGTGTC AGAGGACCTA  
 3081 TGATTATGTC CGGTTATGTA AACAACTCCG AAGCGACCAA CGCCTTGATT GACAAGGATG GATGGCTAC  
 3151 TTCTGGAGAC ATAGCTTACT GGGACGAAGA CGAACACTTC TTCATAGTTG ACCGCTTGAA GTCTTTAATT  
 3221 AAATACAAAG CATACTAGGT GGGCCCCGCT GAATTGGAAT CGATATTGTT ACAACACCCC AACATCTTCG  
 3291 ACGGGGGCGT GGCAAGGTCTT CCCCACGATG ACCGGGGTCA ACTTCCCCC GCGCTTGTG TTTTGGAGCA  
 3361 CGGAAAGACG ATGACGGAAA AAGAGATCGT GGATTACGTC CCCAGTCAAG TAACAACCGC GAAAAGTTG  
 3431 CGCGGAGGAG TTGTGTTGT GGACGAAGTA CCGAAAGGTC TTACCGAAA ACTCGACGCC AGAAAATCA

C<sub>la</sub>I (3259)

3501 CACAGATCCT CATAAAGGCC AAGAAGGGCG GAAAGTCCAA ATTCTAACTC GAGGGGGGGC CCGGTACCTT

Apal (3557)  
XbaI (3548) KpnI (3563)



3571 TAAGACCAAT GACTTACAAG GCAGCTGTAG ATCTTAGCCA CTITTTAAAAA GAAAAGGGGG GACTGGAAGG  
 3641 GCTAATTACAC TCCCCAAAGAA GACAAGATAT CCTTGATCTG TGGATCTACC ACACACAAGG CTACTTCCCT  
 3711 GATTGGCAGA ACTACACACC AGGGCCAGGG GTCAGATATC CACTGACCTT TGGATGGTCC TACAAGCTAG  
 3781 TACCAAGTTGA GCCAGATAAG GTAGAAGAGG CCAATAAAGG AGAGAACACC AGCTTGTAC ACCCTGTGAG  
 3851 CCTGCATGGA ATGGATGACC CTGAGAGAGA AGTGTAGAG TGGAGGTTTG ACAGCCGCCT AGCATTTCAT  
 3921 CACGTGGCCC GAGAGCTGCA TCCGGAGTAC TTCAAGAACT CCTGACATCG AGCTTGCTAC AAGGGACTTT  
 3991 CCGCTGGGGA CTTTCCAGGG AGGCCTGGCC TGGGGGGGAC TGGGGAGTCG CGAGCCCTCA GATGCTGCAT  
 4061 ATAAGCAGCT GCTTTTGCC TGTACTGGGT CTCTCTGGTT AGACCACATC TGAGCCTGGG AGCTCTCTGG  
 4131 CTAACTAGGG AACCCACTGC TTAAGCCTCA ATAAAGCTTGC CTTGAGTGC TTCAAGTAGT GTGTCCCCGT  
 4201 CTGTTGTGTG ACTCTGGTAA CTAGAGATCC CTCAGACCCCT TTTAGTCAGT GTGGAAAATC TCTAGCACCC  
 4271 CCCAGGAGGT AGAGGTTGCA GTGAGCCAAG ATCGGCCAC TGCATTCCAG CCTGGGCAAG AAAACAAGAC  
 4341 TGTCTAAAT AATAATAATA AGTTAAGGGT ATAAATATA TTATACATG GAGGTATAA AAATATATAT  
 4411 ATTTGGGCTG GGGCAGTGG CTCACACCTG CGCCCGGGCC 1TTGGGAGGC CGAGGCAGGT GGATCACCTG  
 4481 AGTTTGGGAG TTCCAGACCA GCCTGACCAA CATGGAGAAA CCCCTCTCT GTGTATTTT ATGAGATTTT  
 4551 ATTTTATGTG TATTTTATTC ACAGGTATT CTGGAAAACCT GAAACTGTT TTCTCTACT CTGATACCAAC  
 4621 AAGAACATCAG AGCACAGAGG AAGACTTCTG TGATCAAATG TGGTGGGAGA GGGAGGTTT CACCAAGCACA  
 4691 TGAGCAGTCAGTCA GTTCTGGCC AGACTCGGGC GGTGCTCTTC GGTTCAGTT CAACACCGCC TGCCCTGGAGA  
 4761 GAGGTCAAGAC CACAGGGTGA GGGCTCAGTC CCCAACACAT AAACACCCAA GACATAAACCA CCCAACAGGT  
 4831 CCACCCCCGCC TGCTGCCAG GCAGAGCCCA TTCACCAAGA CGGGAAATTAG GATAGAGAAA GAGTAAGTCA  
 4901 CACAGAGCCG GCTGTGGGG AGAACGGAGT TCTATTATGA CTCAAATCAG TCTCCCAAG GATTCCGGGA  
 4971 TCAGACTTT TAAGGATAAC TTAGTGTGA GGGGGCCAGT GAGTTGGAGA TGAAGCGTA GGGAGTCCAA  
 5041 GGTGTCTTT TGGCCCGAGT CAGTTCTCG GTGGGGGCCA CAAGATCGGA TGACCCAGTT TATCAATCCG  
 5111 GGGGTGCCAG CTGATCCATG GAGTGCAGGG TCTGCAAAT ATCTCAAGCA CTGATTGATC TTAGTTTTA  
 5181 CAATAGTGAT GTTACCCAG GAACAATTG GGGAAAGGTCA GAATCTTGTG GCCTGTAGCT GCATGACTCC  
 5251 TAAACCATAA TTCTTTTTT GTTTTTTTT TTTTATTTT GACACAGGGT CTCACTCTGT CACCTAGGCT  
 5321 GGAGTGCAGT GGTCAATCA CAGCTCACTG CAGCCCTAG AGCGGCCGCC ACCCCGGTGG AGCTCCAATT  
 5391 CGCCCTATAG TGACTCGTAT TACAATTAC TGGGGCTCGT TTTACAACGT CGTACTGGG AAAACCTGG  
 5461 CGTTACCCAA CTTAATCGCC TTGCAGGCCA TCCCCCTTC GCGAGCTGGC GTAATAGCGA AGAGGCCCGC  
 5531 ACCGATCGCC CTTCCCAACA GTTGCAGGCC CTTGAATGGCG AATGGCGCG AATTTGAAAC GTTAATATTT  
 5601 TGTAAAATT CCCGTTAAAT TTTGTTAAA TCAGCTCATT TTTAACCAA TAGGCCAAA TCGGCAAAAT  
 5671 CCCTTATAAA TCAAAAGAAT AGACCGAGAT AGGGTTGAGT GTTGTTCAGG TTTGGAACAA GAGTCCACTA  
 5741 TTAAAGAACG TGACTCCAA CGTCAAAAGG CGAAAAACCG TCTATCAGG CGATGGCCCA CTACGTGAAC  
 5811 CATCACCCCTA ATCAAGTTT TTGGGGTCCGA GTGCCGTAA AGCAACTAAAT CGGAACCCCTA AAGGGACCCC  
 5881 CGGATTAGA GCTTGACGGG GAAAGCCGGC GAAACGTGGCG AGAAAGGAAG GGAAGAAAGC GAAAGGAGCG

FIG. 10D

5951 GCGGCTAGGG CGCTGGCAAG TGTAGGGTAC CGCTGGCG TAACCACCC ACCGGCCGC CTTAATGCC  
 6021 CGCTACAGGG CGCGTCCCAG GTGGCACTT TCAGGGAAAT CTGGCGGAA CCCCTATTG TTATTTTC  
 6091 TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAT GCTTCAATAA TATTGAAAAA  
 6161 GGAAGAGTAT GAGTATTCAA CATTCCGTG TCGCCCTTAT TCCCTTTT CCGGCATTG GCGCTCCGT  
 6231 TTTGCTCAC CCAGAAAACGC TGCTGAAAGT AAAAGATCCT GAAGATCAGT TGGTGCACG AGTGGGTAC  
 6301 ATCGAACCTG ATCTCAACAG CGCTAACATC CTTCAGACTT TTCCCCCCGA AGAACCTTT CCAATGTGA  
 6371 GCACTTTAA AGTTCTGTA TGTGGCGGG TATTATCCCG TATTGACGCC GGCGAAGAGC AACTCGTCC  
 6441 CGGCATACAC TATTCTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC  
 6511 ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTTCTGA  
 6581 CAACCATCGG AGGACCGAAG GAGCTAACCG CTTTTTGCA AACATGGGG GATCATGTAA CTGGCCTTGA  
 6651 TCGTCCGAA CGCGAGCTGA ATGAAGCCAT ACCAAACGAC GAGCGTGAAC CCACGATGCC TCTAGCAATG  
 6721 GCAACAACTG TGGCCAAATC ATTAACCTGC GAACACTTA CTCTAGCTTC CCGGCAACAA TTAATAGACT  
 6791 GGATGGAGGC GGATAAAAGT CCAGGACAC TCTGCGCTC GGCCCTTCCG GCTGGCTGGT TTATTGCTGA  
 6861 TAAATCTGGA GCCGGTGCAGC GTGGGCTCTG CGGTATCATT GCAGCACTGG GGCCAGATGG TAAGCCCTCC  
 6931 CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAAAG AAATAGACAG ATCGCTGAGA  
 7001 TAGTGGCTC ACTGATTAAAG CATTGGTAAC TGTAGACCCA AGTTTACTCA TATATACCTT AGATTGATTI  
 7071 AAAACTTCAT TTAAATTAA AAAGGATCTA GTGAGATC CTTTTGATA ACTTCATGAC CAAATCCCT  
 7141 TAACGTGAGT TTTCGTTCCA CTGAGCGTCA GACCCCGTAG AAAACATCAA AGGATCTTCT TGAGATCCTT  
 7211 TTTTCTGCG CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTGGCCGGA  
 7281 TCAAGAGCTA CCAACTCTT TTCCGAAGGT AACTGGCTTC AGCAGAGCAG AGATACCAAA TACTGCTCTT  
 7351 CTAGTGTAGC CGTAGTTAGC CCACCACTTC AAGAACTCTG TAGCACCGG TACATACCTC GCTCTGCTAA  
 7421 TCCCTTACCG AGTGGCTGCT GCCAGTGGC ATAAGCTGT TCTTACCGGG TTGGACTCAA GACGATAGT  
 7491 ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGGTTCG TGCACACAGC CACGCTTGGG GCGAACGACC  
 7561 TACACCGAAC TGAGATACCT ACAGCGTCACT CTATGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGCCGG  
 7631 ACAGGTATCC GTTAAGCGGC AGGGTCCGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACCCCTG  
 7701 GTATCTTAT AGTCCCTGTC GGTTTCGCA CCTCTGACTT GAGCGTCGAT TTGTTGATG CTCGTCAAGG  
 7771 GGGCGGAGCC TATGGAAAAAA CGCGCACCAAC CGGGCTTTT TACGGTCTC GCGCTTTTG TGGCCTTTG  
 7841 CTCACATGTT TTTCCTGCG TTATCCCTG ATTCTGTGGA TAACCGTATT ACCGGCTTTG AGTGCACCTGA  
 7911 TACCGCTCGC CGCAGCGGA CGACCCAGCG CAGCGAGTCA GTGAGCGAGG AACCGGAAGA GCGCCCAATA  
 7981 CGCAAACCGC CTCTCCCCCGC CGTTGGCCG ATTCAATTAAAT GCAGCTGGCA CGACAGGTTT CCCGACTGGA  
 8051 AAGCGGGCAG TGAGCGCAAC GCAATTAAATG TGAGTTAGCT CACTCATTAG GCACCCCCAGG CTTTACACTT  
 8121 TATGCTTCCG GCTCGTATGT TGTGTGGAAT TGTCACCGGA TAACAATTTC ACACAGGAAA CAGCTATGAC  
 8191 CATGATTACG CCAAGCTCGG ATTAAACCTT CACTAAAGGG AACAAAAGCT GCTGCAGGGT CCCTAACTGC  
 8261 CAAGCCCCAC AGTGTGCCCT GAGGCTGCCCT TTTCCTTCTA GCGGCTGCC CCACTCGGCT TTGCTTCCC  
 8331 TAGTTCACT TACTTGCCTT CAGCCAAAGT CTGAAACTAG GTGCGCACAG AGCGTAAGA CTGCGAGAGA  
 8401 AAGAGACAG CGTACAGGG GTTTATCAC AGTGCACCC GACAGTCGTC AGCCTCACAG GGGGTTTATC  
 8471 ACATTGCACT CTGACAGTCG TCAGCCTCAC AGGGGGTTTA TCACACTGCA CCCTTACAAT CATTCCATT  
 8541 GATTACAAT TTTTTAGTC TCTACTGTGCT CTAACCTGTG AGTAAATT GATCAGAGGT GTGTTCCAG  
 8611 AGGGGAAAC AGTATATACA CGGTTCACTA CTATCGCATT TCAGGCCCTC ACCTGGGTCT TGGAAATGTGT  
 8681 CCCCCGAGGG GTGATGACTA CCTCAGTTGG ATCTCCACAG GTCAAGTGA CACAAGATAA CCAAGACACC  
 8751 TCCCAAGGCT ACCACAATGG GCGCCCTCC ACGTGCACAT GCGCGGAGGA ACTGCCATGT CGGAGGTCCA  
 8821 AGCACACCTG CGCATCAGAG TCCTTGGTGT GGAGGGAGGG ACCAGCCAG CTCCAGCCA TCCACCTGAT  
 8891 GAACAGAACCG TAGGGAAAGC CCCAGTCTA CTTACACCAG GAAAGGC (SEQUENCE ID NO: 8)

FIG. 10E

1 TGGAAGGGCT AATTGGTCC CAAAAAAGAC AAGAGATCCT TGATCTGTGG ATCTACCACA CACAAGGCTA  
 71 CTTCCCTGAT TGGCAGAACT ACACACCAGG GCCAGGGATC AGATATCCAC TGACCTTTGG ATGGTGCTTC  
 141 AAGTTAGTAC CAGTTGAACC AGAGCAACTA GAAGAGGCCA AATAAGGAGA GAAGAACAGC TTGTTACACC  
 211 CTATGAGCCA GCATGGGATC GAGGACCCGG AGGGAGAACT ATTAGTGTGG AAGTTGACA GCCTCCTAGC  
 281 ATTCGTAC ATGGCCCGAG AGCTGCATCC GGAGTACTAC AAAGACTGCT GACATCGAGC TTTCTACAAG  
 351 GGACTTTCCG CTGGGGACTT TCCAGGGAGG TGTGGCTGG GCGGGACTGG GGAGTCCCGA GCCCTCAGAT  
 421 GCTACATATA AGCAGCTGCT TTTGCCTGT ACTGGGTCTC TCTGCTTAGA CCAGATCTGA GCCTGGAGC  
 491 TCTCTGGCTA ACTAGGGAAC CCACTGCTTA AGCCTCAATA AACCTTGCT TGAGTGCTCA AAGTAGTGTG  
 561 TGCCCCGTCTG TTGTGTGACT CTGGTAACTA GAGATCCCTC AGACCCTTT AGTCAGTGTG GAAAATCTCT  
 631 ACCAGTGGCG CCCGAACAGG GACTTGAAAG CGAAAGTAAA GCCAGAGGAG ATCTCTCGAC GCAGGACTCG  
 701 BssHII (711) GCTTGCTGAA CGCGGCaacgg caagaggcga gggggcgccgC ctgACgagGa cgccaaaaat tttgactagc  
 771 Clai (830) ggaggctaga aggagagagC TCGGTGCGAG ACCGTCAGTA TTAAGCGGGG GAGAATTAGA TCGATGGAA  
 841 AAAATTGGT TAAGGCCAGG GGGAAAGAAG AACTACAAGC TAAAGCACAT CGTATGGCA ACCAGGGAGC  
 911 AccI (959) TAGAACGATT CGCAGTTAAT CCTGGCTGT TAGAAACATC AGAAGGCTGT AGACAAATAC TCGGACACCT  
 981 ACAACCATCC CTTCAGACAG GATCAGAGGA GCTTCGATCA CTATACAACA CAGTAGCAAC CCTCTATTGT  
 1051 GTGCACCAGC GGATCGAGAT CAAGGCACACC AAGGAAGCTT TAGACAAGAT AGACCAAGAG CAAAACAAGT  
 1121 CCAAGAAGAA GGCCCAGCAG GCAGCAGCTG ACACAGGACA CAGCAATCAG GTCAGCCAAA ATTACCCAT

FIG. 11A

1191 AGTGCAGAAC ATCCAGGGC AAATGGTACA TCAGGCCATA TCACCTAGAA CTTTAAACGA TAAGCTTG  
 1261 AGTTCCGGT TACATAACTT ACGGTAAATG GCCCGCCTGG CTGACCGCCC AACGACCCCC CCCATTGAC  
 1331 GTCAATAATG ACCTATGTT CCATAGTAAC GCCAATAGGG ACTTTCATT GACGTCAATG CGTGGAGTAT  
 1401 TTACGGTAAA CTGCCCACCTT CCCAGTACAT CAAGTGTATC ATATGCCAAG TACGCCCCCT ATTGACGTCA  
 1471 ATGACGGTAA ATGGCCCGCC TGGCATTATG CCCACTACAT GACCTTATGG GACTTCCCTA CTTGGCAGTA  
 1541 CATCTACGTA TTAGTCATCG CTATTACCAT GGTGATGGGG TTTTGGCAGT ACATCAATGG CGGTGGATAG  
 1611 CGGTTTGACT CACGGGGATT TCCAAGTCTC CACCCCATG ACGTCAATGG GAGTTTGTIT TGCCACCAAA  
 1681 ATCAACGGGA CTTTCCAAAA TGTGTAACA ACTCCGCCCC ATTGACGCAA ATGGCGGTA GGCGTGTACG  
 1751 GTGGGAGGTC TATATAAGCA GAGCTCGTT AGTGAACCGT CAGATCGCCT GGAGACGCCA TCCACGCTGT  
 1821 TTTGACCTCC ATAGAAGACA CCGACTCTAG AG<sub>gatcc</sub>ATC TAAGTAAGCT TGGCATTCCG GTACTGTTGG  
 1891 TAAAATGGAA GACGCCAAAA ACATAAAGAA AGGGCCGGCG CCATTCTATC CTCTAGAGGA TGGAACCGCT  
 1961 GGAGAGCAAC TGCATAAGGC TATGAAGAGA TACGCCCTGG TTCCCTGGAAC AATTGCTTT ACAGATGCCAC  
 2031 ATATCGAGGT GAACATCACG TACGCCGAAT ACTTCGAAT GTCCGTTGGG TTGGCAGAAG CTATGAAACG  
 2101 ATATGGCCTG AATACAAATC ACAGAATCGT CGTATGCACT GAAAATCTC TTCAATTCTT TATGCCGGTG  
 2171 TTGGCCGGT TATTTATCGG AGTTGCAGTT GCGCCCGCGA ACCACATTAA TAATGAACGT GAAATTGCTCA  
 2241 ACAGTATGAA CATTTCGAG CCTACCGTAG TGTTTGTTC CAAAAAGGGG TTGCAAAAAA TTTTGAACGT  
 2311 GCAAAAAAAA TTACCAATAA TCCAGAAAAT TATTATCATG GATTCTAAAA CGGATTACCA GGGATTTCAG  
 2381 TCGATGTACA CGTTCGTCAC ATCTCATCTA CCTCCCGGTT TTAATGAATA CGATTTGTAA CCAGAGTCCT

2451 TTGATCGTGA CAAAACAATT GCACTGATAA TGAATTCCCTC TGGATCTACT GGGTTACCTA AGGGTGTGGC  
 2521 CTTCCGCAT AGAACTGCCT GCGTCAGATT CTGGCATGCC ACAGATCCTA TTTTGCAA TCAAATCATT  
 2591 CGGGATACTG CGATTTAAG TGTGTTCCA TTCCATCACG GTTTGGAAT GTTTACTACA CTCGGATATT  
 2661 TGATATGTGG ATTCGAGTC GTCTTAATGT ATAGATTGA AGAAGAGCTG TTTTACGAT CCCTTCAGGA  
 2731 TTACAAAATT CAAAGTCGT TGCTAGTACC AACCTATTT TCATTCTCG CCAAAAGCAC TCTGATTGAC  
 2801 AAATACGATT TATCTAATT ACACGAAATT GCTTCTGGGG GCGCACCTCT TTGAAAGAA GTCGGGGAAG  
 2871 CGGTTGAAA ACGCTTCAT CTTCCAGGGA TACGACAAGG ATATGGCTC ACTGAGACTA CATCAGCTAT  
 2941 TCTGATTACA CCCGAGGGGG ATGATAAACCG GGGCGGGTC GGTAAAGTTG TTCCATTGTT TGAAGCGAAG  
 3011 GTTGTGGATC TGGATACCGG GAAAACGCTG GGCGTTAAC AGAGAGGCGA ATTATGTGTC AGAGGACCTA  
 3081 TGATTATGTC CGGTTATGTA ACAATCCGG AAGCGACCAA CCCCTTGATT GACAAGGATC GATGGCTACA  
 3151 TTCTGGAGAC ATAGCTTAATC GGGACGAAGA CGAACACTTC TTCATAGTTG ACCGCTTGAA GTCTTTAATT  
 3221 AAATACAAAG GATATCAGGT GGCCTCCGCT GAATTGGAAT CGATATTGTT ACAACACCCC AACATCTTCG  
 3291 ACGCGGGCGT GCCAGGTCTT CCCGACCGATG ACGCCGGTGA ATTCCTCCGCC GCGTTGTTG TTTGGAGCA  
 3361 CGGAAAGACG ATGACGGAAA AAGAGATCGT GGATTACGTC GCCAGTCAG TAACAACCGC GAAAAAGTTG  
 3431 CGCGGAGGAG TTGTGTTGT GGACGAAGTA CCGAAAGGTC TTACCGGAAA ACTCGACGCA AGAAAAATCA

Clal (3259)

3501 GAGAGATCCT CATAAAGGCC AAGAAGGGCG GAAAGTCCAA ATTGTAAcTC GAGGGGGGGC CCGGTACCTT

ApaI (3557)  
XhoI (3548) KpnI (3563)

3571 TAAGACCAAT GACTTACAAG GCAGCTGTAG ATCTTAGCCA CTTTTAAAAA GAAAAGGGGG CACTGGAAGG

3641 GCTAATTAC TCCCAAAGAA GACAAGATAT CCTTGATCTG TGGATCTACC ACACACAAGG CTACTTCCCT  
 3711 GATTGCCAGA ACTACACACC AGGGCCAGGG GTCAGATATC CACTGACCTT TGGATGGTGC TACAAGCTAG  
 3781 TACCACTTGA GCCAGATAAG GTAGAACAGG CCAATAAAGG AGAGAACACC AGCTTGTAC ACCCTGTGAG  
 3851 CCTGCATGGA ATGGATGACC CTGAGAGAGA AGTGTAGAG TGGAGGTTG ACAGCCGCCT AGCATTTCAT  
 3921 CACGTGGCCC GAGAGCTGCA TCCGGACTAC TTCAAGAACT GCTGACATCG AGCTTGTAC AAGGGACTTT  
 3991 CCCGCTGGGA CTTTCCAGGG AGCCGTGCC TGGGGGGAC TGGGGACTGG CGAGCCCTCA GATGCTGCAT  
 4061 ATAAGCAGCT GCTTTTGCC TGTACTGGT CTCTCTGGT AGACCAGATC TGAGCCTGG AGCTCTCTGG  
 4131 CTAACTAGGG AACCCACTGC TTAAGCCTCA ATAAAGCTT CTTTGAGTGC TTCAAGTAGT GTGTGCCCGT  
 4201 CTGTTGTGTG ACTCTGGTAA CTAGAGATCC CTCAGACCCCT TTTAGTCAGT GTGGAAAATC TCTAGCACCC  
 4271 CCCAGGAGGT AGAGGTTGCA GTGAGCCAAG ATCCGCCAC TGCATTCCAG CCTGGGCAAG AAAACAAGAC  
 4341 TGCTAAAAAT AATAATAATA AGTTAAGGT ATTAATATA TTTATACATG GAGGTATAA AAATATATAT  
 4411 ATTGGGCTG GGGCAGTGG CTCACACCTG CGCCCGGCC TTTGGAGGC CGAGGAGGT GGATCACCTG  
 4481 AGTTGGGAG TTCCAGACCA GCCTGACCAA CATGGAGAAA CCCCTTCTCT GTGTATTTT AGTAGATTIT  
 4551 ATTTTATGTG TATTTTATTC ACAGGTATTT CTGGAAAATC GAAACTGTTT TTCCCTACT CTGATACAC  
 4621 AAGAACATC AGCACAGAGG AAGACTCTG TGATCAAATG TGGTGGGAGA GGGAGGTTT CACCAGCACA  
 4691 TGAGCAGTCA GTTCTGCCG AGACTCGGG GGTGCTCTTC GTTCACTTC CAACACCGCC TGCCTGGAGA  
 4761 GAGGTAGAC CACAGGGTGA GGGCTCAGTC CCCAAGACAT AAACACCCAA ACATAAACAC CCAACACAGT  
 4831 CCACCCGCC TGCTGCCAG GCAGAGCCGA TTCAACAAAGA CGGGAAATTAG GATAGAGAAA GAGTAAGTCA  
 4901 CACAGAGCCG GCTGTGGGG AGAACGGACT TCTATTATGA CTCAAATCAG TCTCCCCAAG CATTGGGGA  
 4971 TCAGAGTTT TAAGGATAAC TTAGTGTGA GGGGGCCAGT GAGTTGGAGA TGAAAGCGTA GGGAGTCGAA  
 5041 CCTGTCCTT TGGCGCCGAGT CAGTTCTGG GTGGGGGCCA CAAGATCGGA TGAGCCAGTT TATCAATCCG  
 5111 GGGGTGCCAG CTGATCCATG GAGTCCAGGG TCTCAAAAT ATCTCAAGCA CTGATTGATC TTAGGTTTAA  
 5181 CAATAGTGTG TTACCCCAAG GAACAATTG GGGAAAGGTCA GAAATTTGTA GCCTGTAGCT CCATGACTCC  
 5251 TAAACCATAA TTCTTTTTT GTTTTTTTT TTTTATTTT GAGACAGGGT CTCACCTGT CACCTAGGCT  
 5321 GGAGTGCAGT GGTGCAATCA CAGCTCACTG CAGCCCTAG AGCGGGCCGC ACCGGGGTGG AGCTCCAATT  
 5391 CGCCCTATAG TGAGTCGTAT TACAATTAC TGGCCGTGTT TTTACAACTG CGTGAETGGG AAAACCTCG  
 5461 CGTTACCCAA CTAAATCGCC TTGCAAGACA TCCCCCTTC GCCAGCTGGC GTAATAGCGA AGAGGGCCCG  
 5531 ACCGATCGCC CTTCACCAACA GTTGCAGCAGC CTGAATGGCG AATGGCCCA ATTTCAAAC TTAAATATT  
 5601 TGTTAAAATT CGCGTTAAAT TTTGTAAA TCAAGCTATT TTTAACAA TAGGCCAAA TCGGCAAAAT  
 5671 CCCTTATAAA TCAAAAGAAT AGACCGAGAT AGGGTTGAGT GTTGTTCAG TTTGGAACAA GAGTCCACTA  
 5741 TTAAAGAACG TGGAATCCAA CGTCAAAGGG CGAAAACCG TCTATCAGGG CGATGGCCCA CTACGTGAAC  
 5811 CATCACCTA ATCAAGTTT TTGGGGTGA GTGCCCTAA AGCACTAAAT CGGAACCTA AAGGGAGCCC  
 5881 CCGATTAGA GCTTGACGGG GAAACCCGGC GAAAGTGGCG AGAAAGGAAG GGAAGAAAGC GAAAGGAGCCG  
 5951 GGCGCTAGGG CGCTGCCAAC TGTAGCGTC ACCGCTGCCG TAACCACCA ACCGGCCGG CTTAATGCGC  
 6021 CGCTACAGGG CGCGTCCCGC GTGGCAGTTT TCCGGAAAT GTGGCGGGAA CCCCTATTG TTATTTTTC  
 6091 TAAATACATT CAAATATGTA TCCGCTCATG AGACAATAAC CCTGATAAA GCTTCAATAA TATTGAAAAA

FIG. 11D

6161 CGAACAGTAT GAGTATTCAA CATTTCGGTG TCCCCCTTAT TCCCTTTTTT CCGGCATTTT CCCTTCCITG  
 6231 TTTTGCTCAC CCAGAAACGC TGTTGAAAGT AAAAGATGCT GAAGATCAGT TGGGTGCACG AGTGGGTTAC  
 6301 ATCGAAGTGG ATCTCAACAG CGGTAAGATC CTTGAGAGTT TTGCCCCGA AGAACGTTT CCAATGATGA  
 6371 GCACCTTAA AGTTCTGCTA TGTTGGCGGG TATTATCCCG TATTGACGCC GGGCAAGAGC AACTCGGTGCG  
 6441 CCGCATAACAC TATTCTCAGA ATGACTTGGT TGACTACTCA CCAGTCACAG AAAAGCATCT TACGGATGCC  
 6511 ATGACAGTAA GAGAATTATG CAGTGCTGCC ATAACCATGA GTGATAACAC TGCAGGCAAC TTACTCTGA  
 6581 CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTGCA AAACATGGGG CATCATGTAAC CTGGCCTTGA  
 6651 TCCTGGGAA CCCGGAGCTGA ATGAAGCCAT ACCAACCGAC GAGCGTGACA CCACCATGCC TGTAGCAATG  
 6721 GCAACAACTG TGCGCAAACCT ATTAACTGGC GAACTACTTA CTCTACCTTC CGGGCAACAA TTAATAGACT  
 6791 GGATGGAGGC GGATAAAAGTT GCAGGACCAC TTCTGCGCTC GGGCCCTTCCG GCTGGCTGGT TTATTGCTGA  
 6861 TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCACTT GCAGCACTGG GCCCAAGATGG TAAGCCCTCC  
 6931 CGTATCGTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA  
 7001 TAGGTGCTC ACTGATTAAG CATTGTAAC TGTCAAGCCA AGTTTACTCA TATATACTTT AGATTGATTT  
 7071 AAAACTTCAT TTTTAATTAA AAAGGATCTA GGTGAAGATC CTTTTGATA ATCTCATGAC CAAAATCCT  
 7141 TAACGTGAGT TTTCCTTCCA CTGAGCGTCA GACCCCGTAG AAAAGATCAA AGGATCTCT TGAGATCCTT  
 7211 TTTTCTGCG CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCGCTACCA GCGGTGGTTT GTTGGCCGGA  
 7281 TCAAGAGCTA CCAACTCTT TTCCGAAGGT AACTGGCTC ACCAGAGGCG AGATACAAA TACTGTCCTT  
 7351 CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACCTG TAGCACCCGC TACATACCTC CCTCTGCTAA  
 7421 TCCTGTTACC AGTGGCTGCT GCGAGTGGC ATAAGTCTG TCTTACCGGG TTGGACTCAA GACGATAGTT  
 7491 ACCGGATAAG CGCGAGCGGT CGGGCTGAAAC GGGGGGTTCG TGACACACG CCAGCTTGGG CGAACGACC  
 7561 TACACCGAAC TGAGATAACCT ACAGCGTGAAC CTATGAGAAA CCCCCACGCT TCCCGAAGGG AGAAAGGCGG  
 7631 ACAGGTATCC GTAAAGCGGC AGGGTCGAA CAGGAGAGCG CACGAGGGAG CTTCCAGGGG GAAACGCCCTG  
 7701 GTATCTTAT AGTCTGTGCG CGTTTCCCA CCTCTGACTT GAGCGTCGAT TTTTGTGATC CTGGTCAGGG  
 7771 GGGCGGAGCC TATGGAAAAA CGCCAGCAAC GGGGCTTCA TACGGTTCTT GGCCTTTGC TGGCCTTTTG  
 7841 CTCACATGTT CTTTCTGCG TTATCCCCCTG ATTCTGTGGA TAACCGTATT ACCGCTTTC AGTGAGCTGA  
 7911 TACCGCTCCG CGCAGCGGAA CGACCGAGCG CACCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCCAATA  
 7981 CGCAAACCGC CTCTCCCCGC GCGTTGGCCG ATTCTTAAT GCAGCTGGCA CGACAGGTTT CCCGACTGGA  
 8051 AAGCGGCGAG TGAGCGCAAC GCAATTAAAT TGAGTTAGCT CACTCTATTAG GCACCCCAAGG CTTTACACTT  
 8121 TATGCTTCCG GCTCGTATGT TGTGCGGAAT TGTGAGCGGA TAACAAATT ACACAGGAAA CAGCTATGAC  
 8191 CATGATTAGC CCAAGCTCGG AATTAAACCT CACTAAGGG AACAAAAGCT GCTGCAGGGT CCCTTAACCTG  
 8261 CAAGCCCCAC AGTGTGCCCT GAGGCTGCCCTT CTTCTTCTA GCGGCTGCC CCACTCGGCT TTGCTTCCC  
 8331 TAGTTCACT TACCTGGTT CAGCCAAGGT CTGAAACTAG GTGGCAGACAG AGCGGTAAAGA CTGGAGAGA  
 8401 AAGAGACAG CTTTACAGGG GTTTTATCAC AGTGCACCCCT GACAGTCGTC ACCCTCACAG GGGGTITATC  
 8471 ACATTGCCACC CTGACAGTCG TCAGCCTCAC AGGGGGTTA TCACAGTGCA CCCTTACAAT CATTCCTATT  
 8541 GATTCAACAT TTTTTAGTC TCTACTGTGCT CTAACCTGTA AGTTAAATT GATCAGAGGT GTGTTCCAG  
 8611 AGGGAAAAC AGTATATACA GGGTTCACTA CTATCGCATT TCAGGCCCTC ACCTGGGTCT TGGAAATGTGT  
 8681 CCCCCGAGGG GTGATGACTA CCTCACTTGG ATCTCCACAG GTCACAGTGA CACAAGATAA CCAAGACACC  
 8751 TCCCAAGGCT ACCACAATGG GCGGCCCTCC ACGTGCACAT GGCCGGAGGA ACTGCCATGT CGGAGGTGCA  
 8821 AGCACACCTG CGCATCAGAG TCCTTGGTGT GGAGGGAGGG ACCAGGCCAG CTTCCAGCCA TCCACCTGAT  
 8891 GAACAGAACCG TAGGGAAAGC CCCAGTCTA CTTACACCAG GAAAGGC (SEQUENCE ID NO: 9)

FIG. 11E

mBCwCN frag ----- C - AC - G -----  
m2BCwCN frag ----- C - G - G -----  
BC/HXB2 -----  
BC/NL43 -----  
#1 .....  
CGCGCACGGC AAGAGGCGAG GGGCGCCGAC TGGTGAGTAC GCCAAAAATT

mBCwCN frag ----- C - C -----  
m2BCwCN frag -----  
BC/HXB2 ----- T -----  
BC/NL43 ----- G -----  
#51 .....  
TTGACTAGCG GAGGCTAGAA GGAGAGAGAT GGGTGCGAGA GCGTCAGTAT

mBCwCN frag -----  
m2BCwCN frag -----  
BC/HXB2 -----  
BC/NL43 ----- AA -----  
#101 .....  
TAAGCGGGGG AGAATTAGAT CG

FIG. 12

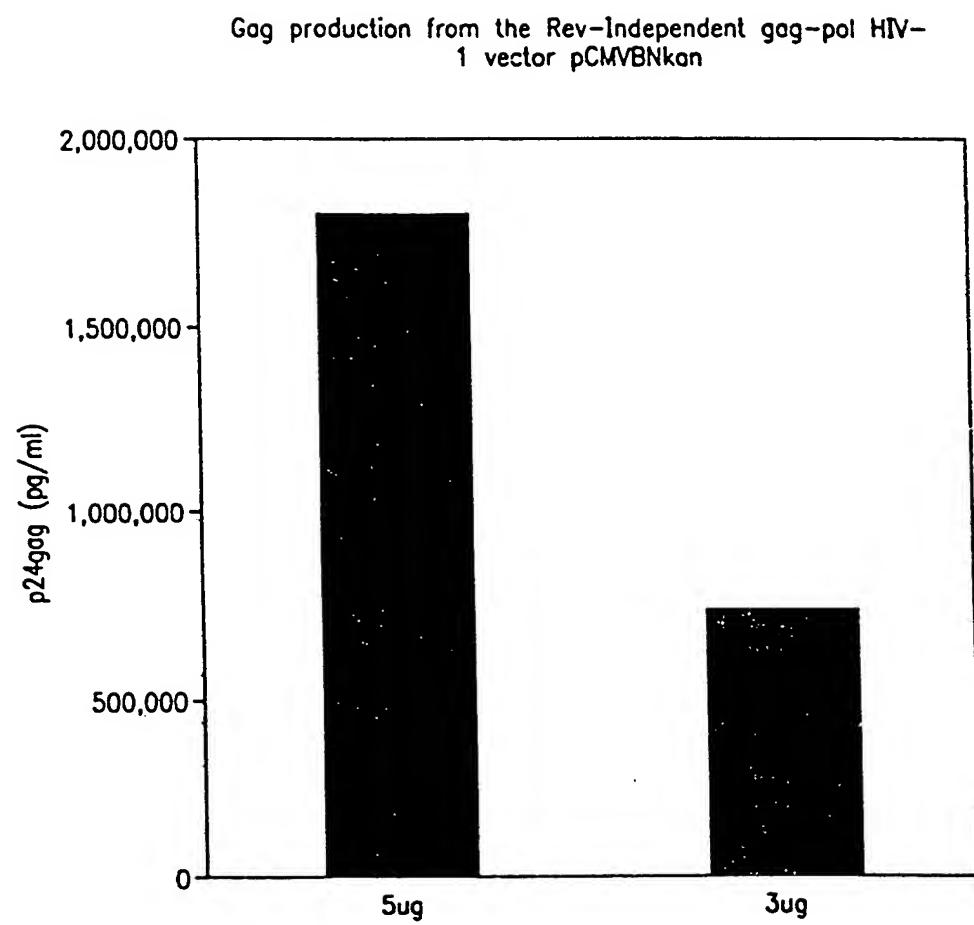


FIG. 13

Reverse transcriptase activity from the Rev-  
Independent gag-pol HIV-1 vector pCMVBNkan

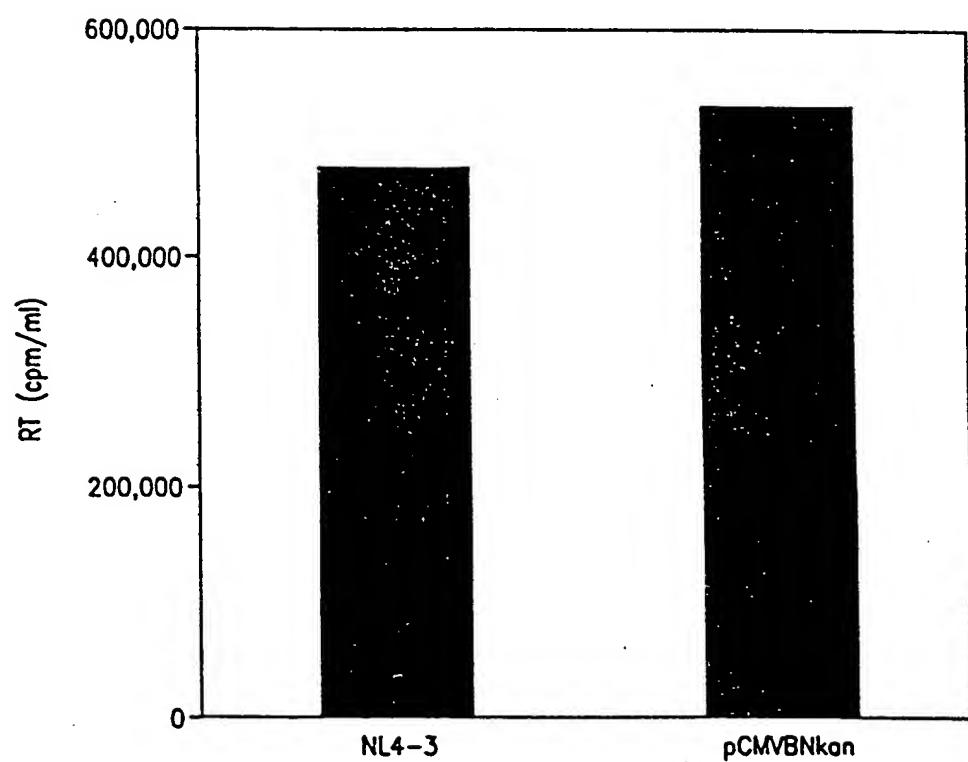


FIG. 14

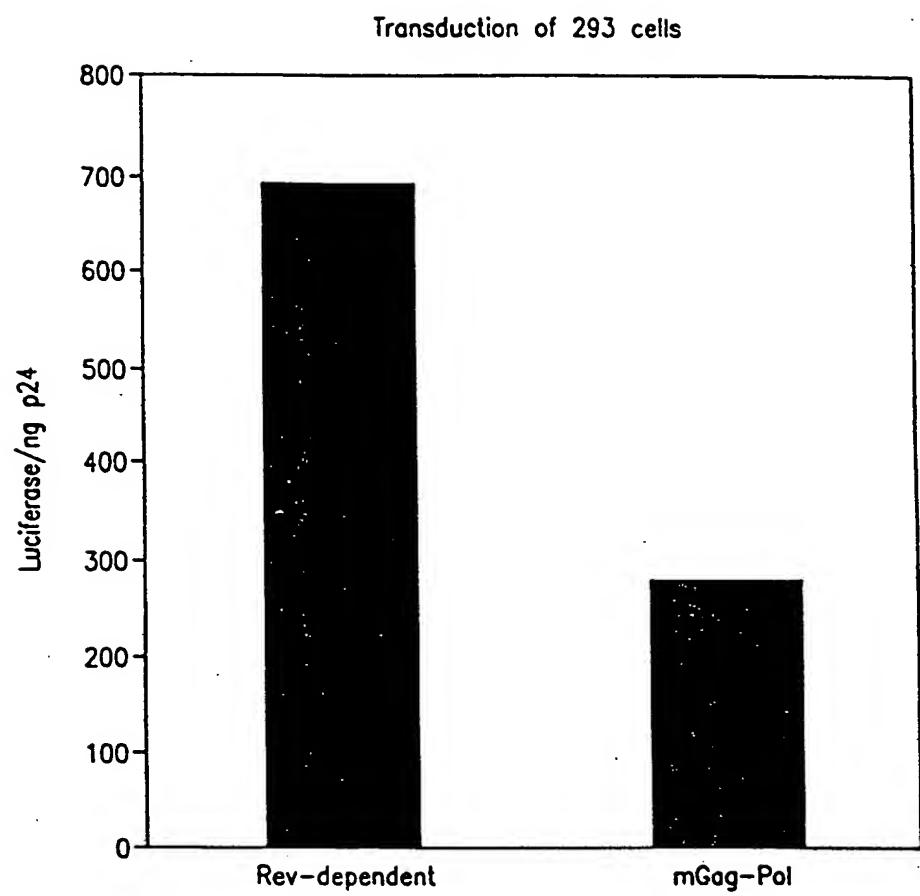


FIG. 15A

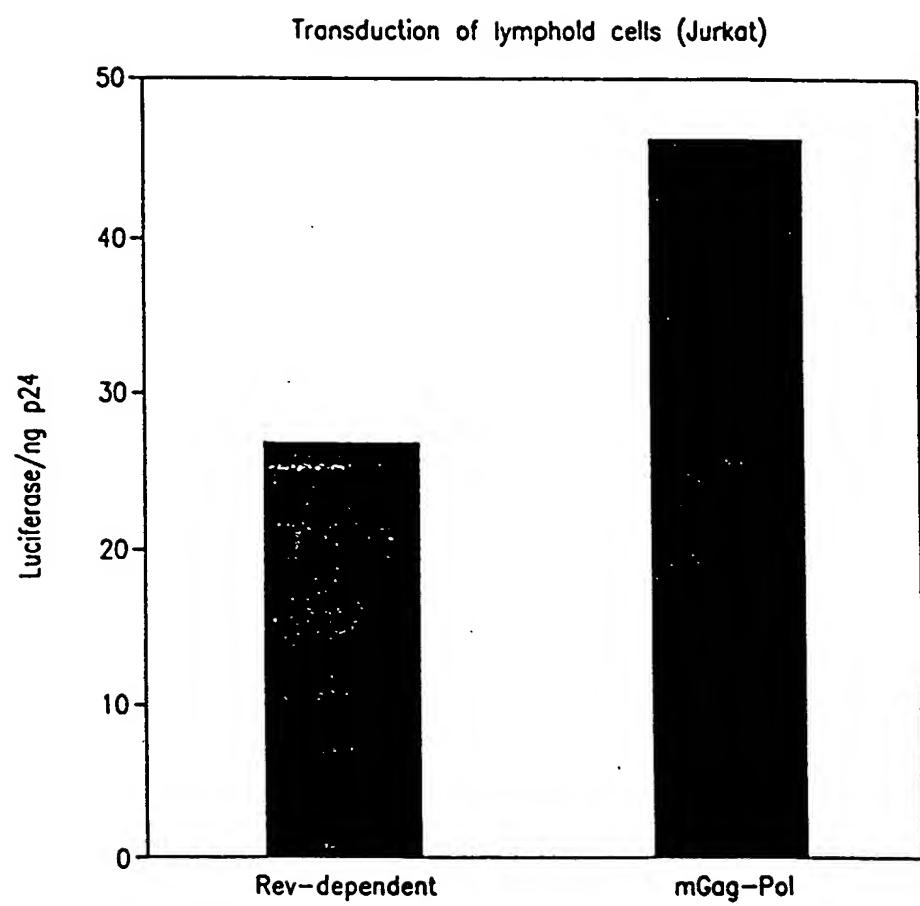


FIG. 15B

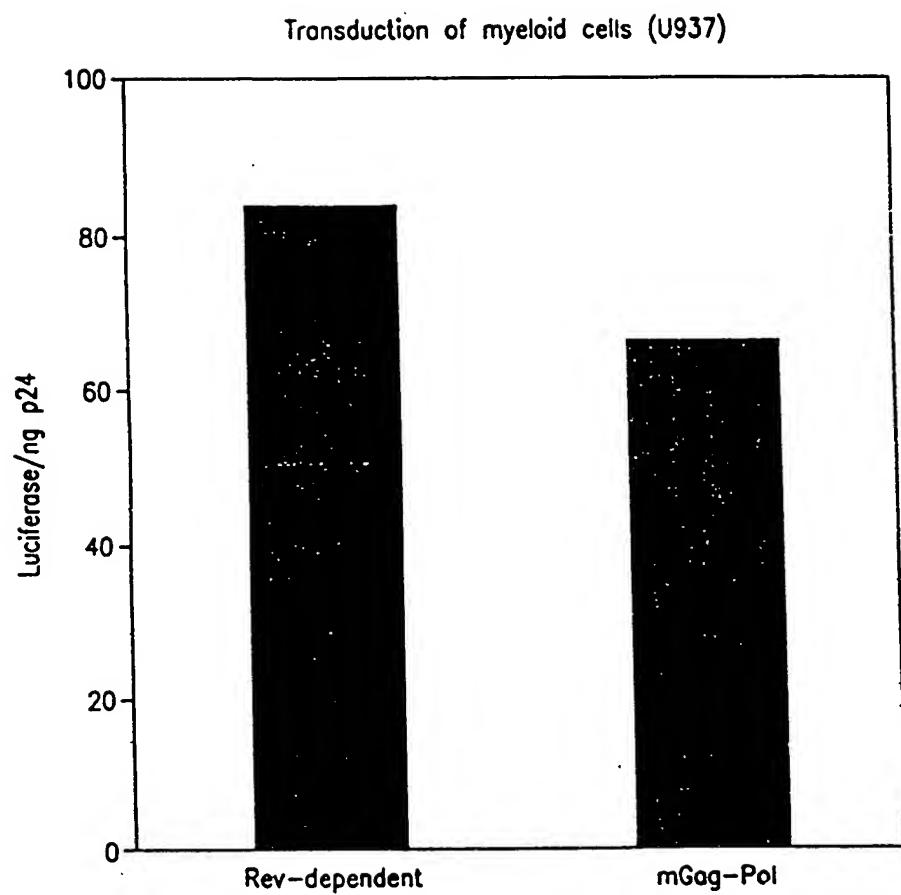


FIG. I5C

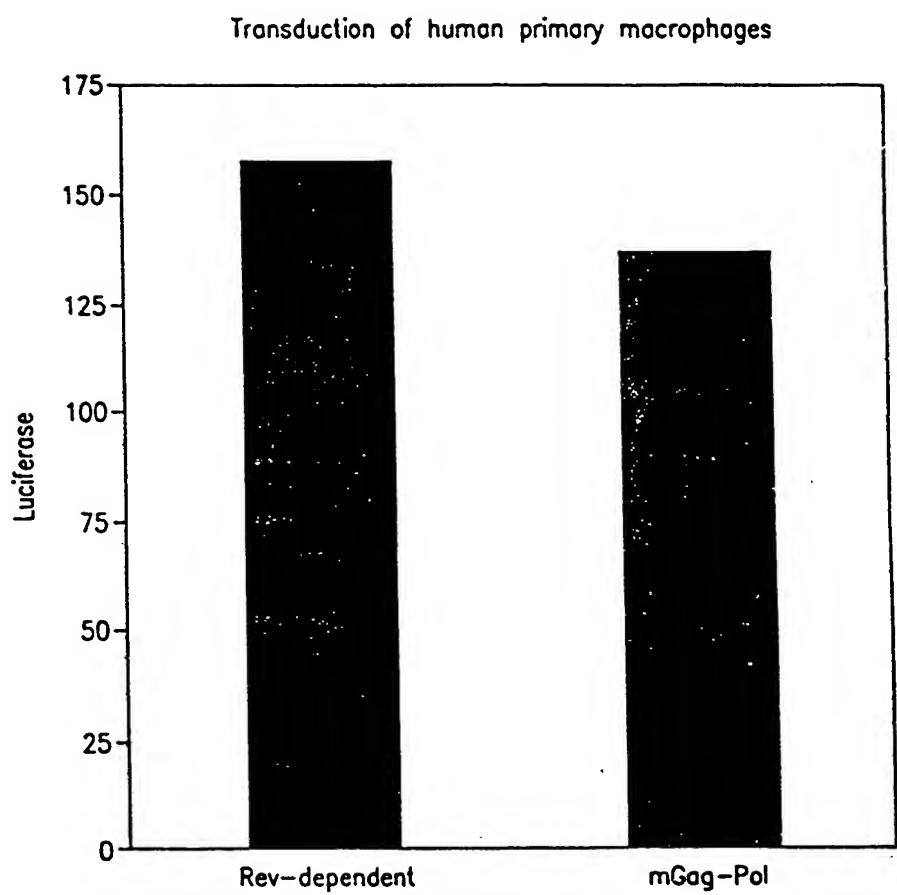


FIG. 15D

41/45

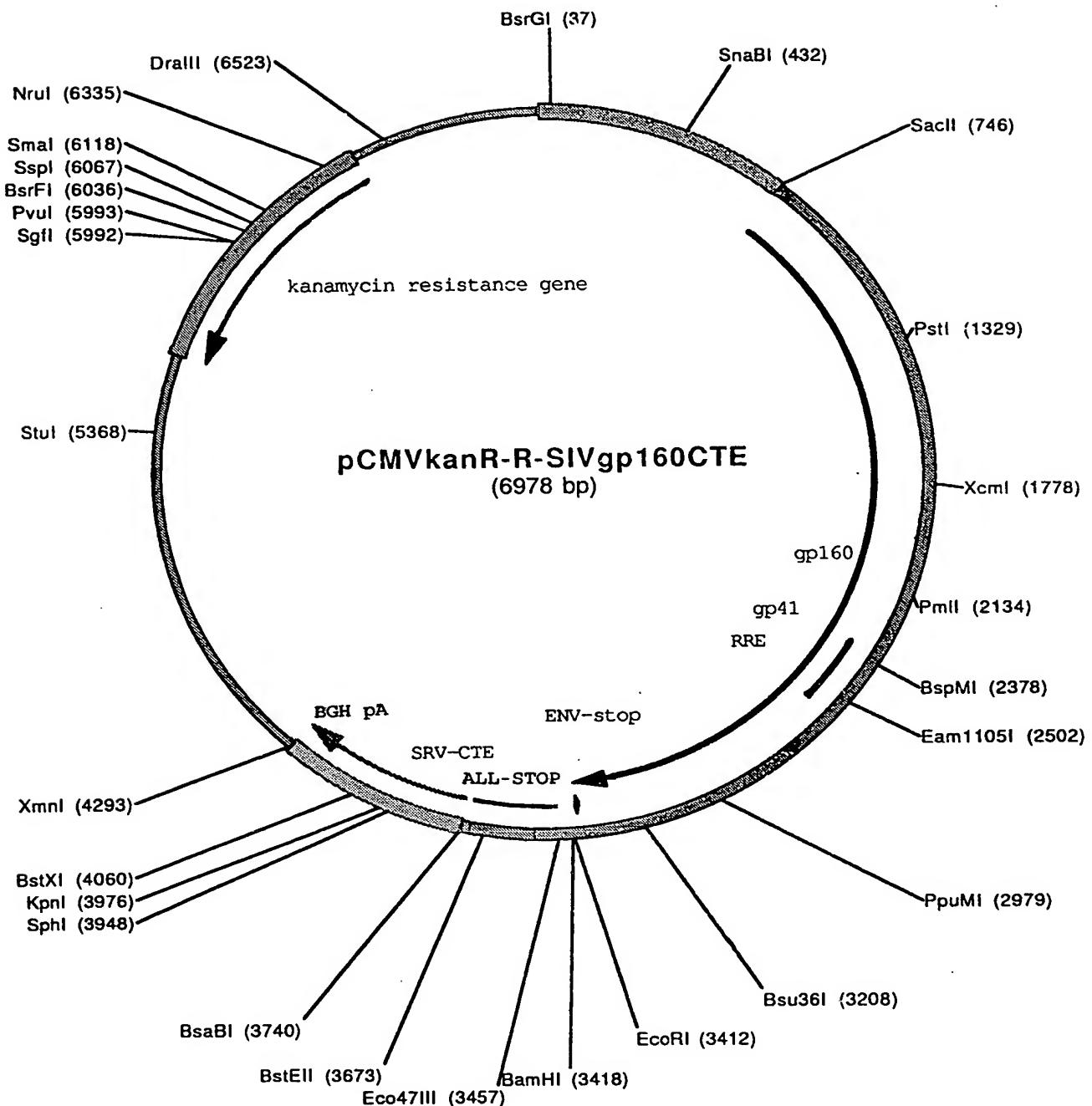


FIG. 16

# 42/45

BsrGI (37)

1 CCTGGCCATTGCATACGTTGTATCCATATCATAATAATGTCACATTATIGGCTCATGTCACACATTACGCCATGTTGA  
 81 CATTGATTATTGACTAGTTATTAAATAGTAATCAATTACGGGGCATTAGTICATAGCCCATATATGGAGTTCCCGCTTAC  
 161 ATAACCTACGGTAAATGGCCGCTGGTGCACCCCCAACGACCCCCCCTATTGACGTCATAATGACGTATGTTCCA  
 241 TAGTAACGCAATAGGACTTCCATTGACGTCATAATGGTGGAGTATTACGGTAACTGCCACTTGGCAGTACATCAA  
 321 GTGTATCATATGCCAAGTAGCCTTCTATTGACGTCATAATGCCAGTACATGAC

SnaBI (432)

401 CTTATGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGTATTACCATGGTATGCGGTTGGCAGTACA  
 481 TCAATGGGGTGGATAACGGGTTGACTCACGGGATTCCAAGTCTCCACCCCCATTGACGTCATAATGGAGTTGGTTGG  
 561 CACCAAATCAACGGACTTCCAAAATGTCGTAACAACCTCGCCCCATTGACGCAATGGCGGTACGGTGTACGGTG  
 641 GGAGGTCTATATAAGCAGAGCTCGTTAGTGAACGTCAGATGCCCTGGAGACGCCATCCACGCTGTTGACCTCCATA

SacII (746)

721 GAAGACACCGGGACCGATCCAGCCTCCGGCGCTAAAGTATGGATGTCTGGGAATCAGCTGCTATGCCATCT

1► Met Gl yCysLeuGl yAsnGl nLeuLeuIleAl aIleAl

801 TGCTTTAAAGTGTCTATGGATCTATTGACTCTATATGTCACAGTCTTATGGTGTACAGCTGGAGGAATGCCACAG

13► euLeuLeuSer Val TyrGl yIleTyrCysThr LeuTyrVal Thr Val PheTyrGl yVal ProAl aTrpArgAsnAl aThr  
 881 ATTCCCCCTTTTGCAACCAAGAATAGGGATACTTGGGAACAACTCAGTGCCTACAGATAATGGTGTATTTCAGA

40► IleProLeuPheCysAl aThr LysAsnArgAspThr TrpGl yThr Thr Gl nCysLeuProAspAsnGl yAspTyrSer Gl  
 961 AGTGGCCCITAATGTTACAGAAAGCTTGTGCTGGATAATAACAGTCACAGAACAGGAATAGAGGATGTATGGCAAC

66► uValAl aLeuAsnVal Thr Gl uSer PheAspAl aTrpAsnAsnThr Val Thr Gl uGl nAl aIleGl uAspVal TrpGl nL  
 1041 TCTTGAGACCTCAATAAAGCCTTGTGAAAATTATCCCCATTATGCATTACTATGAGATGCAATAAGTAGAGACAGAT

93► euPheGl uThr Ser IleLysProCysVal LysLeuSer ProLeuCysIleThr Met ArgCysAsnLysSer Gl uThrAsp  
 1121 AGATGGGATTGACAAATCAATAACAAACAGCATCAACACATCAACGACAGCATCAGCAAAGTAGACATGGTCAA

120► ArgTrpGl yLeuThr LysSer IleThr Thr Thr Al aSer Thr Thr Ser Thr Thr Al aSer Al aLysVal AspMet Val As  
 1201 TGAGACTAGTTCTGTATAGCCCAGGATAATTGCACAGGCTGGACAAGAGCAAATGATAAGCTGTAAATTCAACATGA

146► pGl uThr Ser Ser CysIleAl aGl nAspAsnCysThr Gl yLeuGl uGl nGl uGl nMet IleSer CysLysPheAsnMet T  
 1281 CAGGGTTAAAAGAGACAAGAAAAAGAGTACAATGAAACTTGGTACTCTGCAGATTGGTATGTGAACAAGGAATAAC

173► hr Gl yLeuLysArgAspLysLysGl uTyrAsnGl uThr TrpTyrSer Al aAspLeuVal CysGl uGl nGl yAsnAsn  
 1361 ACTGGTAATGAAAGTAGATGTTACATGAACCACTGTAACACTCTGTATCAAGAGCTGTGACAAACATTATGGGA

200► Thr Gl yAsnGl uSer ArgCysTyrMetAsnHisCysAsnThr Ser Val IleGl nGl uSer CysAspLysHisTyrTrpAs  
 1441 TGCTATTAGATTAGGTATTGTGCACCTCCAGGTATGCTTAGATGACACAAATTATCAGGCTTATGC

226► pAl aIleArgPheArgTyrCysAl aProProGl yTyrAl aLeuLeuArgCysAsnAspThrAsnTyrSer Gl yPheMet P  
 1521 CTAAATGTTCTAAGGTGGTCTCTCATGCACAAGGATGGAGACACAGACTCTACTTGGTTGGCTTAAATGGA

253► r oLysCysSer LysVal Val Val Ser Ser CysThr ArgMetMet Gl uThr Gl nThr Ser Thr TrpPheGl yPheAsnGl y  
 1601 ACTAGAGCAGAAAATAGAACCTTATTTACTGGCATGGTAGGGATAATAGGACTATAATTAGTTAAATAAGTATATAA

280► Thr ArgAl aGl uAsnArgThr TyrIleTyrTrpHisGl yArgAspAsnArgThr IleIleSer LeuAsnLysTyrTyrAs  
 1681 TCTAACATGAAATGAGAACAGGAAATAAGACAGTTTACCACTATGCTGGATTGGTTTCCACTC

306► nLeuThrMetLysCysArgArgProGl yAsnLysThr Val LeuProVal Thr IleMetSer Gl yLeuVal PheHisSer G  
 XcmI (1778)

1761 AACCAATCAATGATAGGCCAAGCAGGCATGGTGTGGTTGGAGGAAATGGAAGGATGCAATAAGAGGTGAAGCAG

333► InProIleAsnAspArgProLysGl nAl aTrpCysTrpPheGl yGl yLysTrpLysAspAl aIleLysGl uVal LysGl n  
 1841 ACCATGTCACACATCCCAAGGTAACTGGAACAAACTGATAAAATCAATTGACGGCTCTGGAGGAGGAGATCC

360► Thr IleVal LysHisProArgTyrThr Gl yThrAsnAsnThrAspLysIleAsnLeuThr Al aProGl yGl yAspPr  
 1921 CGAAGTTACCTCATGTGGACAAATTGCAAGAGAGTCTCTACTGTAAAATGAATTGGTTCTAAATTGGTAGAAG

386► oGl uVal Thr PheMetTrpThrAsnCysArgGl yGl uPheLeuTyrCysLysMetAsnTrpPheLeuAsnTrpVal Gl uA  
 2001 ATAGGAATACAGCTAACCGAACAGCAAGAACAGCATAAAAGGAATTACGTGCCATGTCATAATTAGACAAATACTAAC

413► spArgAsnThr Al aAsnGl nLysProLysGl uGl nHisLysArgAsnTyrVal ProCysHis II ArgGl nIleIleAsn

# FIG. 17

Pml (2134)

2081 ACTTGGCATAAAAGTAGGCAAAATGTTATTGCCTCCAAGAGAGGGAGACCTCACGTGTAACCCACAGTGACCAGTCT  
 440► Thr Trp His Lys Val Gl y Lys Asn Val Tyr Leu Pro Pro Arg Gl u Gl y Asp Leu Thr Cys Asn Ser Thr Val Thr Ser Le  
 2161 CATAGCAAACATAGATGGATTGATGGAAACCAAACAAATATCACCATGAGTCAGAGGTGCCAGAAGTGTATCGATTGG  
 466► Ile Al a Asn Ile Asp Trp Ile Asp Gl y Asn Gl n Thr Asn Ile Thr Met Ser Al a Gl u Val Al a Gl u Leu Tyr Arg Leu G  
 2241 AATTGGGAGATTATAAATTAGTAGAGATCACTCCAATTGGCTGGCCCCACAGATGTGAAGAGGTACACTACTGGTGGC  
 493► IuLeuGl yAspTyrLysLeuVal Gl u IleThr Pro IleGl yLeuAl aProThrAspVal LysArgTyrThr Thr Gl yGl y  
 BspMI (2378)  
 2321 ACCTCAAGAAATAAAAGAGGGCTTGTGCTAGGGTTCTGGGTTCTCGAACGGCAGGTCTGCAATGGGAGCCGC  
 520► Thr Ser Arg Asn Lys Arg Gl y Val Phe Val Leu Gl y Phe Leu Gl y Phe Leu Al a Thr Al a Gl y Ser Al a Met Gl y Al a Al  
 2401 CAGCCTGACCCCTCACGGCACAGTCCCACACTTATTGGCTGGGATAGTCCAACAGCAGCAACAGCTGGACGTGGTCA  
 546► aSer Leu Thr Leu Thr Al a Gl n Ser Arg Thr Leu Leu Al a Gl y Ile Val Gl n Gl n Gl n Gl n Leu Leu Asp Val Val L  
 Eam 1105I (2502)  
 2481 AGAGACAACAAGAATTGGCGACTGACCGTCTGGGAAACAAAGAACCTCCAGACTAGGGTCACTGCCATCGAGAAAGTAC  
 573► ys Arg Gl n Gl n Gl u Leu Leu Arg Leu Thr Val Trp Gl y Thr Lys Asn Leu Gl n Thr Arg Val Thr Al a Ile Gl u Lys Tyr  
 2561 TTAAAGGACCAGCGCAGCTGAATGCTTGGGATGTGCGTTAGACAAGTCTGCCACACTACTGTACCATGGCAAATGC  
 600► Leu Lys Asp Gl n Al a Gl n Leu Asn Al a Trp Gl y Cys Al a Phe Arg Gl n Val Cys His Thr Thr Val Pro Trp Pro Asn Al  
 2641 AAGCTAACACAAAGTGGAACATGAGACTTGGCAAGAGTGGAGCGAAAGGTGACTCTTGGAGAAAATAAACG  
 626► aSer Leu Thr Pro Lys Trp Asn Asn Gl u Thr Trp Gl n Gl u Trp Gl u Arg Lys Val Asp Phe Leu Gl u Gl u Asn Ile Thr A  
 2721 CCCTCCTAGAGGAGGCACAAATTCAACAAGAGAACATGTATGAATTACAAAAGTTGAATAGCTGGATGTGTTGGC  
 653► IaLeuLeuGl u Gl u Al a Gl n Ile Gl n Gl u Lys Asn Met Ty Gl u Leu Gl n Lys Leu Asn Ser Trp Asp Val Phe Gl y  
 2801 AATTGGTTGACCTTGCTCTGGATAAAAGTATATAATGGAGTTATATAGTTGAGTAATACTGTTAAGAAT  
 680► Asn Trp Phe Asp Leu Al a Ser Trp Ile Lys Tyr Ile Gl n Tyr Gl y Val Tyr Ile Val Val Gl y Val Ile Leu Leu Arg II  
 2881 AGTGATCTATAGTACAAATGCTAGCTAACGTTAAGGCAGGGTATAGGCCAGTGTCTCTCCCCACCCCTTTATTCC  
 706► eVal Ile Tyr Ile Val Gl n Met Leu Al a Lys Leu Arg Gl n Gl y Tyr Arg Pro Val Phe Ser Ser Pro Pro Ser Tyr Phe G  
 PpuMI (2979)  
 2961 AGCAGACCCATATCCAACAGGACCCGGCACTGCCAACAGAGAAGGCAAAGAGACGGTGGAGAAGCGGTGGCAAC  
 733► I n Gl n Thr His I l e Gl n Gl n Asp Pro Al a Leu Pro Thr Arg Gl u Gl y Lys Gl u Arg Asp Gl y Gl y Gl u Gl y Gl y Asn  
 3041 AGCTCCTGGCCTGGCAGATAGAATATATCCACTTTCTATTGCTAGCTTATTGACTCTTGGCTATTAGTAA  
 760► Ser Ser Trp Pro Trp Gl n Ile Gl u Tyr Ile His Phe Leu Ile Arg Gl n Leu Ile Arg Leu Leu Thr Trp Leu Phe Ser As  
 3121 CTGTAGGACTTTGCTATCGAGACTATACAGATCTCCAACCAACTCCAGGCTCTGCGACCCCTACAGAGGATC  
 786► nCysArgThr Leu Leu Ser Arg Val Tyr Gl n Ile Leu Gl n Pro Ile Leu Gl n Arg Leu Ser Al a Thr Leu Gl n Arg Ile A  
 Bsu36I (3208)  
 3201 GAGAAGTCCCTCAGGACTGAAC TGACCTACCTACAATATGGGTGAGCTATTCCATGAGGCCGTCAGCCGCTGGAGA  
 813► r g Gl u Val Leu Arg Thr Gl u Leu Thr Tyr Leu Gl n Tyr Gl y Trp Ser Tyr Phe His Gl u Al a Val Gl n Al a Val T r p Arg  
 3281 TCTGCGACAGAGACTCTGGGGCGGTGGGAGACTTATGGGAGACTCTAGGAGAGGTGGAAGATGGATACTCGCAAT  
 840► Ser Al a Thr Gl u Thr Leu Al a Gl y Al a Trp Gl y Asp Leu Trp Gl u Thr Leu Arg Arg Gl y Gl y Arg Trp Ile Leu Al a I  
 BamHI (3418)  
 3361 CCCCAGGAGGATTAGACAAGGGCTTGAGCTCACTCTTGTGAGGGACAGAGAATTGGATCCactagtctagaCTOGA  
 866► eProArg Arg Ile Arg Gl n Gl y Leu Gl u Leu Thr Leu Leu . . .  
 Eco47III (3457)  
 3441 GGGGGGGGCCGGTACGAGCGCTTAGCTAGCTAGAGACCACCTOCCCTGCGAGCTAACGCTGGACAGCCAA TGACGGTAAG  
 3521 AGAGTGACATTTCACTAACCTAACAGACAGGAGGGCCGTAGAGCTACTGCCATATCAAAGACGGTAAAAGTGATAAA  
 3601 AATGTATCACTCCAACCTAACAGACAGGCGCAGCTCCGAGGGATTGCGTCTGTTATATATATTTAAAAGGGTGACCT  
 BstEII (3673)

# 44/45

BsaBI (3740)

3681 GTCCGGAGCCGTGCTGCCCGATGATGTCCTGGCTAGACTCGAGGGGGGCCGGTACGATCCAGATCTGCTGCCCTT

3761 CTAGTTGCCAGCCATCTGTTGCCCCCTCCCCCGTGCCCTCCTGACCCCTGGAAGGTGCCACTCCCACTGTCCTTCC

3841 TAATAAAATGAGGAAATTGCATCGCATTGTCAGTAGGTCTCATTCTATTCTGGGGGGTGGGTGGGGCAGCACAGCAA

SphI (3948)

KpnI (3976)

3921 GGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGATGCGGTGGGCTATGGGTACCCAGGTGCTGAAGAATTGAC

BstXI (4060)

4001 CCGGTTCCCTGGGCCAGAAAGAACGAGCAGGCCATCCCCTCTGTGACACACCCCTGTCACGCCCTGGTTCTTAGTT

4081 CCAGCCCCACTCATAGGACACTCATAGCTCAGGAGGGCTCCGCTTCAATCCACCCGCTAAAGTACTTGGAGCGGTCTC

4161 TCCCTCCCTCATCAGCCCCACCAACCAAACCTAGCCTCCAAGAGTGGGAAGAAATTAAAGCAAGATAGGCTATTAAAGTGC

XbaII (4293)

4241 AGAGGGAGAGAAAATGCCTCCAACATGTGAGGAAGTAATGAGAGAAAATCATAGAATTCTCCGCTTCCCTCGCTCACTGA

4321 CTCGCTCGCTCGGTGTCGGCTGCGCGAGCGGTATCAGCTCACTCAAAGCGGTAAACGGTTATCCACAGAACATCAG

4401 GGGAATAACGCAGGAAAAGAACATGTGAGCAAAGGCCAGAACGCGTAAAAGGCCCGTTGCTGGGTTT

4481 TTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCAGCCTCAAGTCAGGAGTGGCAACCCGACAGGACTATA

4561 AAGATCAGGCGTTCCTCCCTGAGGCTCCCTGTCAGGCTCCTCTGTCAGGCTTACCTGGCTTACCCGATAACCTGGCTTACCCG

4641 CCTTTCTCCCTCGGAAAGCGTGGGCTTCTCAATGTCAGGCTGTAGGTATCTCAGTTGGTGTAGGTGTTCCGCTCC

4721 AAGCTGGCTGTGTCACGAACCCCCCGTTCAGCCGAGCCACTGGTAACAGGATTAGCAGAGGAGGTATGTAAGGCGGTACAG

4801 GGTAAGACACGACTTATGCCACTGGCAGCCACTGGTAACAGGATTAGCAGAGGAGGTATGTAAGGCGGTACAG

4881 AGTTCTGAAGTGGTGGCTAACTACGGCTACACTAGAAGGACAGTATTGGTATCTGCGCTGCTGAAGCCAGTACAG

4961 TTCGAAAAAGAGTTGGTAGCTCTGATCCGCAAACAAACCCGCTGGTAGCGGTGTTTTTTGTTGCAAGCAGCA

5041 GATTAACGGCAGAAAAAGGATCTAACAGAAGATCTTCTACCGGGTCTGACGCTCAGTGGAAACGAAACACT

5121 CACGTTAAAGGGATTGGTAGGATTATCAAAAGGATCTCACCTAGATCCTTTAAATTAAAAAGTAAAGTTTAAA

5201 TCAATCTAAAGTATATGAGTAAACTTGGTCTGACAGTTAACGCTTAACTCAGTGGCCTATCTCAGCAGTCG

5281 TCTATTGTTCATCCATAGTTGCCCTGACTCGGGGGGGGGCGCTGAGGTCTGCCCTCGTGAAGAAGGTGTTGCTGAC

StuI (5368)

5361 TCATACCAGGCCTGAATGCCCATCATCAGCCAGAAAGTGGGGAGCCACGGTGTAGAGGCTTGTAGGTGGA

5441 CCAGTTGGTGAATTGAACCTTGGCTTGGCACGGAACGGTCTGCGTTGCGGAAGATGCGTGTATGATCCCTCACT

5521 CAGCAAAAGTCGATTATTCAACAAAGCCCGTCCCGTCAAGTCAGGCTAATGCTTCCGCTAACGCTAACACCAATTAA

5601 CCAATTCTGATTAGGAAACACTCATGAGCATCAAATGAAACTGCAATTATTATCATATCAGGATTATCAATACCATATT

271 PhePheGluAspLeuMetLeuHisPheGluLeuLysAsnMetAspProAsnAspIleGlyTyrLysG

5681 TGAAAAAGCCGTTCTGTAATGAAGGAGAAAACCTACCGAGGCAGTTCCATAGGATGGCAAGTCGGTATCGGTCTGC

248 1 nPheLeuArgLysGluLeuSerProSerPheGluGlyLeuCysAsnTrpLeuIleAlaLeuAspGlnTyrArgAspAla

5761 GATTCCGACTCGTCAACATCAATACAACCTATTAAATTCCCTCGTAAAAATAAGGTTATCAAGTGGAAACATCACC

222 1 IleGlyValArgGlyValAspIleCysGlyIleLeuLysGlyGluAspPhenIleLeuAsnAspLeuSerPheAspGlyHi

5841 GAGTGACGACTGAATCGGTGAGAAATGCAAAGCTTATGCAATTCTCCAGACTTGTCAACAGGCCAGCCATTACGC

195 1 sThrValValSerAspProSerPheProLeuLeuLysHimMetGluLeuLysTrpValGlyInGluValProTrpGlyAsnArgG

PvuI (5993)

SgfI (5992)

5921 TCGTCATCAAATCACTCGCATCAACCAAACCGTTATTCACTCGTATTGCGCTGAGCGAGCGAAATACGCGATCGCT

168 1 IuAspAspPheAspSerAlaAspValIleLeuGlyAsnAsnMetArgSerGlnAlaGlnAlaLeuArgPheValArgAspSer

BsrFI (6036)

SspI (6067)

6001 GTTAAAAGGACAATTACAAACAGGAATCGAACCGGGCAGGAACACTGCCAGGCCATCAACATAATTTCACCTG

142 1 AsnPheProCysAsnCysValProIleSerHisLeuArgArgLeuPheValAlaLeuAlaAspValIleAsnGluGlySer

SmaI (6118)

6081 AATCAGGATATTCTCTAAACCTGGATGCTGTTCCGGGATCGCAGTGGTAGTAACCATGCATCATCAGGAGTA

115 1 rAspProTyrGluLeuValGlyInPheAlaThrLysGlyProIleAlaThrThrLeuLeuTrpAlaAspAspProThrA

6161 CGGATAAAATGCTGATGGCGGAAGAGGCAATAATTCCGTCAGCCAGTTAGTCTGACCATCTCATCTGAAACATCATT

88 1 gIlePheHiLysIleThrProLeuProMetPheGluThrLeuTrpAsnLeuArgValMetGluAspThrValAspAsn

6241 GGCACGCTACCTTGCATGTTICAGAAACAACTCTGCCATCGGGCTTCCCATACAAATCGATAAGTTGTCGACCTG

6241 62 1 AlaValSerGlyLysLeuPheLeuGluProAlaAspProLysGlyTyrLeuArgTyrIleThrAlaGlySer

NruI (6335)

6321 ATIGCCCGACATTATCCGAGGCCATTATACCCATATAAACATCAGCATCCATGTTGGAATTAAATCCGGCCTCGAGCAA

35 1 rGlnGlyValAsnAspArgAlaTrpLysTyrGlyTyrLeuAspAlaAspMetAsnSerAsnLeuArgProArgSerCysS

6401 GACGTTCCCGTTGAATATGGCTCATAACACCCCTTGATTACTGTTATGTAAGCAGACAGTTATTGTCATGATGA

84 1 rThrGluArgGlyInIleHiLysSerMet

DraIII (6523)

6481 TATATTTTATCTTGCAATGTAACATCAGAGATTTTGAGACACAAACGTTGGCTTCCCCCCCCCATTATTGAAAGCA

6561 TTTATCAGGGTTATTGTCATGAGCGGATACATATTGAATGTTAGAAAATAACAAATAGGGTCCGCGCACA

**45/45**

6641 TTTCCCCGAAAAGGCCACCTGACGTCAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCCGTATCACGAG  
6721 GCCCTTCGCTCGCGCTTTCGGTGTGACGGTAAAAACCTCTGACACATGCAGCTCCCGAGACGGTCACAGCTTGTC  
6801 TGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTGGCGGGTGTGGCTTAACCTAT  
6881 GCGGCATCAGAGCAGATTGACTGAGAGTGACCATATGCGGTGTGAAATACCGACAGATGCCAAGAGAAAATACCG  
6961 CATCAGATTGGCTATTGG